

**Predictive models of pregnancy based on data from a preconception cohort study**

**Running Title: Predictive models of pregnancy**

Jennifer J Yland<sup>a,\*</sup>; Taiyao Wang<sup>b,c,\*</sup>; Zahra Zad<sup>b,d,\*</sup>; Sydney K Willis<sup>a</sup>; Tanran R Wang<sup>a</sup>; Amelia  
K Wesselink<sup>a</sup>; Tammy Jiang<sup>a</sup>; Elizabeth E Hatch<sup>a</sup>; Lauren A Wise<sup>a</sup>; Ioannis Ch. Paschalidis<sup>b,d,e</sup>

<sup>a</sup> Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA;

<sup>b</sup> Center for Information and Systems Engineering, Boston University, Boston, MA, USA;

<sup>c</sup> Philips Research North America, Cambridge, MA, USA;

<sup>d</sup> Department of Electrical and Computer Engineering, Division of Systems Engineering, Boston  
University, Boston, MA, USA;

<sup>e</sup> Department of Biomedical Engineering, Boston University, Boston, MA, USA;

\* These authors contributed equally to this work and should be considered joint first authors.

**Corresponding Author:**

Jennifer J Yland

yland@bu.edu

(ORCID 0000-0001-7870-8971)

Department of Epidemiology

Boston University School of Public Health

715 Albany Street

23 Boston, Massachusetts, 02118, USA

## 24 **Abstract**

25 **Study question:** Can we derive adequate models to predict the probability of conception among  
26 couples actively planning a pregnancy?

27 **Summary answer:** Leveraging data collected from female participants in a North American  
28 preconception cohort study, we developed models to predict pregnancy with performance  
29 ranging from 63.8% to 71.2% in Receiver Operating Characteristic Area Under the Curve  
30 (AUC).

31 **What is known already:** Earlier work has focused primarily on identifying individual risk  
32 factors for infertility. Several predictive models have been developed in subfertile populations,  
33 with relatively low discrimination (AUC: 59-64%).

34 **Study design, size, duration:** Study participants were female, aged 21-45 years, residents of the  
35 United States or Canada, not using fertility treatment, and actively trying to conceive at  
36 enrollment (2013-2019). Participants completed a baseline questionnaire at enrollment and  
37 follow-up questionnaires every two months for up to 12 months or until conception. We used  
38 data from 4,133 participants with no more than one menstrual cycle of pregnancy attempt at  
39 study entry.

40 **Participants/materials, setting, methods:** On the baseline questionnaire, participants reported  
41 data on sociodemographic factors, lifestyle and behavioral factors, diet quality, medical history,  
42 and selected male partner characteristics. 163 predictors were considered in this study. We  
43 implemented regularized logistic regression, support vector machines, neural networks, and  
44 gradient boosted decision trees to derive models predicting the probability of pregnancy: (i) in  
45 fewer than 12 menstrual cycles of pregnancy attempt time (Model I), and (ii) within six menstrual  
46 cycles of pregnancy attempt time (Model II). Cox models were used to predict the probability of

pregnancy within each menstrual cycle for up to 12 cycles of follow-up (Model III). We assessed model performance using the AUC and the weighted-F1 score for Models I and II, and the concordance index for Model III.

**Main results and the role of chance:** Model I and II AUCs were 70% and 66%, respectively, in parsimonious models, and the concordance index for Model III was 63%. The predictors that were positively associated with pregnancy in all models were having previously breastfed an infant and use of multivitamins or folic acid supplements. The predictors that were inversely associated with pregnancy in all models were female age, female BMI, and history of infertility. Among nulligravid women with no history of infertility, the most important predictors were female age, female BMI, male BMI, menstrual cycle irregularity, use of a fertility app, attempt time at study entry, and perceived stress.

**Limitations, reasons for caution:** Reliance on self-reported predictor data could have introduced misclassification, which would likely be non-differential with respect to the outcome. There may have also been misclassification of time-to-pregnancy, which was estimated using self-reported data on menstrual cycle length and date of the last menstrual period. Further, we cannot be certain that all relevant predictor variables were considered. Finally, though we validated the models using split sample replication techniques, we were unable to conduct an external validation study.

**Wider implications of the findings:** Given a wide range of predictor data, machine learning algorithms can be leveraged to analyze epidemiologic data and predict the probability of conception with discrimination that exceeds earlier work.

**Study funding/competing interest(s):** The research was partially supported by the U.S. National Science Foundation and the National Institutes for Health. In the last three years, Dr.

70 Wise has received in-kind donations for PRESTO research from FertilityFriend.com,  
71 Kindara.com, Sandstone Diagnostics, and Swiss Precision Diagnostics. Dr. Wise also serves as a  
72 fibroid consultant to AbbVie, Inc. The other authors declare no competing interests.

73

74 **Keywords:** fertility, fecundability, pregnancy, prospective studies, predictive analytics, machine  
75 learning.

76

## Introduction

Ten to fifteen percent of couples in North America experience infertility, defined as the inability to conceive within 12 months of regular unprotected intercourse (Thoma et al., 2013). In the U.S., up to 12% of reproductive aged women and 9.4% of men aged 25–44 years used fertility treatments in 2006–2010 (Chandra et al., 2013). The costs of these services exceed \$5 billion in the U.S. annually (Macaluso et al., 2010) and are expected to increase as couples delay childbearing. Developing better prognostic tools for couples trying to conceive could inform clinical care and mitigate potential costs. For women who are concerned about their fertility potential before they start trying to conceive, an accurate predictive model could facilitate decisions about how long to delay childbearing or how to prioritize other potentially modifiable factors.

Previous research has identified many individual risk factors for infertility and predictors of fecundability (*i.e.*, the per-cycle probability of conception). Female age and body mass index (BMI), as well as male BMI, have been identified as risk factors for infertility (Best and Bhattacharya, 2015, Homan et al., 2007, Sundaram et al., 2017, Wesselink et al., 2017). In addition, female preconception exposures including alcohol consumption (Fan et al., 2017); sleep quality (Willis et al., 2019); cigarette smoking (Wesselink et al., 2019); use of certain hormonal contraceptives (Yland et al., 2020); dietary factors (Gaskins and Chavarro, 2018); depressive symptoms (Evans-Hoeker et al., 2018, Nillni et al., 2016); stress (Akhter et al., 2016, Louis et al., 2011, Lynch et al., 2014, Wesselink et al., 2018); and environmental exposures such as air pollution (Conforti et al., 2018) and endocrine disrupting chemicals (Kahn et al., 2020) are associated with reduced fecundability. Other male risk factors include exposure to environmental

chemicals (Buck Louis et al., 2016, Snijder et al., 2012), cigarette smoking (Soares and Melo, 2008), and short sleep duration (Wise et al., 2018). However, few studies have moved beyond individual risk factors to develop predictive models of pregnancy probability, and the predictive power of these models was modest (Coppus et al., 2009).

In this study, we used supervised machine learning methods to predict the cumulative probability of pregnancy over six and 12 menstrual cycles and to predict fecundability (per-cycle probability of conception) in an incident cohort study of pregnancy planners. We considered 163 potential predictors and applied several classification algorithms and variable selection procedures to identify the most accurate models and to evaluate the relative predictive strength of individual risk factors.

## **Methods**

### ***Study population***

Pregnancy Study Online (PRESTO) is a web-based preconception cohort study designed to examine the extent to which lifestyle and behavioral factors such as diet, exercise, and medication use influence fertility and pregnancy outcomes (Wise et al., 2015). The study began in 2013 and is ongoing. Eligible female participants are aged 21-45 years, residing in the U.S. or Canada, trying to conceive, and not using fertility treatments. We excluded participants with more than one menstrual cycle of pregnancy attempt time at enrollment because these women may have changed their behaviors in response to difficulties conceiving (Joffe et al., 2005, Wise et al., 2020). This study included data from 4,133 participants enrolled during 2013 through 2019.

## ***Data collection***

Female participants completed a baseline questionnaire at enrollment, on which they reported data on sociodemographic factors, lifestyle and behavioral factors, medical history, and selected male partner characteristics. Ten days after enrollment, participants were invited to complete the diet history questionnaire (DHQ II). The DHQ II was designed by the National Cancer Institute and the first version of the DHQ was validated against 24-hour dietary recalls in a U.S. population (Millen et al., 2006, Subar et al., 2001). In validation studies, correlations between energy-adjusted, DHQ-reported food servings and 24-hour recall-reported food servings ranged from 0.43 for other starchy vegetables to 0.84 for milk. Based on dietary factors reported via the DHQ II, we assessed overall diet quality using the Healthy Eating Index-2010 (HEI-2010) score (Guenther et al., 2013). Participants completed bimonthly follow-up questionnaires for 12 months, or until reported pregnancy, cessation of pregnancy attempts, study withdrawal, or loss to follow-up, whichever occurred first. Data on menstrual cycle dates, pregnancy attempts, and pregnancy status were obtained via the baseline questionnaire and updated on each follow-up questionnaire. A complete list of the 163 variables included in this analysis is provided in Table I.

## ***Outcomes***

We developed three models to predict 1) pregnancy in fewer than 12 menstrual cycles; 2) pregnancy within six menstrual cycles; and 3) the average probability of pregnancy per menstrual cycle. We chose these outcome measures to reflect clinically relevant definitions of infertility, subfertility, and fecundability (Evers, 2002, Gnoth et al., 2005). For the first two

models, we used a dataset with one observation per participant and excluded participants who were lost to follow-up before reaching a study endpoint (for the first model,  $N = 3,195$ ; for the second model,  $N = 3,476$ ). For the third model (fecundability), we included all participants under observation regardless of follow-up duration ( $N = 4,133$ ).

### ***Pre-processing***

We performed several data reduction steps to prepare the dataset for feature selection and to avoid model overfitting (Hawkins, 2004). First, we converted categorical variables into indicator variables and standardized each predictor by subtracting its mean and dividing by its standard deviation. Next, for each pair of highly correlated variables (correlation coefficient  $>0.8$ ), we removed the variable that had a lower correlation with the outcome to avoid issues of collinearity. We then evaluated the difference in means or proportions between participants with and without a pregnancy, using the chi-squared test (Cochran, 1952) for binary predictors and the Kolmogorov-Smirnov test for continuous predictors (Massey, 1951). We removed the variables that were not significantly associated with the outcome ( $p > 0.05$ ).

### ***Classification methods***

We compared four supervised classification methods to develop predictive models for pregnancy (Hastie et al., 2009, Jiang et al., 2020). Supervised machine learning is an approach in which a dataset is randomly split into a training dataset and a testing dataset. Then, an algorithm (described in greater detail below) is applied to the training data to infer a function that maps a combination of inputs (*i.e.*, predictors) to outputs (*i.e.*, the outcome pregnancy). In a process called feature selection (or elimination), the predictive ability of the model is optimized by

selecting variables to improve prediction of the outcome. This is carried through a series of iterations where models are trained on part of the training set and evaluated on a hold-out training subset (*i.e.*, cross-validation). The model with the final selected set of variables is then trained on the entire training set and its performance evaluated on the testing set.

For Models I (pregnancy in less than 12 menstrual cycles) and II (pregnancy within six menstrual cycles), we first fit logistic regression models with an added regularization term to penalize an overfit model (Friedman et al., 2010). Models derived using regularization are robust to the presence of outliers in the training data set (Chen and Paschalidis, 2018, Chen et al., 2019). We considered both an  $\ell_1$ -norm (L1LR) and an  $\ell_2$ -norm regularizer (L2LR) (Lee et al., 2006). The former is appropriate if we believe that few variables are predictive of the outcome (sparse model), whereas the latter is appropriate in cases where a dense model is more appropriate. Second, we used Support Vector Machines (SVMs), which find a separating hyperplane in the variable space so that the data points from the two different classes reside on different sides of that hyperplane (Cortes and Vapnik, 1995). We considered both a standard linear SVM (L2SVM) and a linear SVM with an  $\ell_1$ -norm regularizer (L1SVM) designed to induce a sparse solution. Third, we used the Light Gradient Boosting Machine (LightGBM) algorithm, which is an ensemble tree-based model that uses a gradient boosting framework (Friedman, 2002, Mason et al., 1999). Fourth, we used artificial neural networks (ANNs), which attempt to organize data based on structures inspired by mammalian brain functioning (Ripley, 2007). As the brain is comprised of neurons, ANNs process data through nodes, which are in turn aggregated into layers. We used Feed Forward Multilayer Perceptron Neural Networks (MLP), with at least three layers of nodes (an input layer, a hidden layer, and an output layer) (Salcedo-Sanz, 2016). In a

feed-forward ANN, information moves in one direction: from input to output. Because there are intermediate layers of information, the MLP algorithm can model complex non-linear relationships. We tuned several hyperparameters including the number of hidden neurons, the number of layers, and the number of iterations. For training, we used a Rectified Linear Unit (ReLU) activation function for the hidden layer and applied the “Adam” optimizer (Kingma and Ba, 2014). These algorithms were chosen because of their extensive usage, interpretability of the output, and performance superiority demonstrated in the literature (Brisimi et al., 2018, Hao et al., 2020, Wang et al., 2020).

We present results for full, sparse, and parsimonious models. The full models (*i.e.*, least parsimonious) contain all variables selected after statistical feature selection (eliminating variables with no statistically significant relationship with the outcome). The sparse models contain variables selected after both statistical feature selection and recursive feature elimination. Recursive feature elimination is a feature selection algorithm that ranks the predictors selected into the full model by importance and iteratively eliminates the least important variables, ultimately selecting a small set of variables that maximize the Area Under the Receiver Operating Characteristic Curve (AUC) in the testing dataset. The parsimonious models were generated by limiting recursive feature elimination to select a model with up to 15 variables. Specifically, we used L1LR to obtain weights associated with the coefficients of the model and eliminated the variable with the smallest absolute weight. We then performed L1LR to obtain a new model and repeated this process until the final model was selected. The final model maximizes a metric equal to the mean AUC minus the standard deviation (std) of the AUC in the testing dataset (described in more detail below). The parsimonious models are easier to

implement and interpret relative to the full and sparse models, which have more variables but similar discrimination. To accommodate categorical variables that were recoded as indicator variables in the preprocessing phase, we selected a reference level for each categorical variable and forced every non-reference level to be included in a model if any other (non-reference) level of the categorical variable was selected.

For Model III (fecundability), we fit a discrete-time analog of the Cox proportional hazards model with cycle number as the time scale, allowing for delayed entry into the risk set (*i.e.*, if a participant already had one cycle of pregnancy attempt at enrollment). Participants contributed at-risk cycles to the analysis from enrollment until reported pregnancy or a censoring event, which included initiation of fertility treatment, withdrawal from the study, cessation of pregnancy attempts, loss-to-follow-up, or 12 cycles of pregnancy attempt, whichever occurred first. We present results for the full model after statistical feature selection, as described above, and for a parsimonious model. To derive the parsimonious model, we fit separate Cox models with each individual predictor and then sorted the variables based on each model's concordance index. The concordance index is similar to the AUC (described below) but accounts for event time and loss to follow-up (Longato et al., 2020, Schmid et al., 2016). We selected the top fifteen variables and forced non-selected levels of polytomous categorical variables into the final model, as described above.

### ***Sensitivity analysis***

We repeated our analyses restricting to nulligravid women with no history of infertility to evaluate the robustness of our results in a population that was presumably naïve to their fertility status.

### ***Performance metrics***

For Models I and II, we primarily evaluated model performance using the AUC. The AUC, or C-statistic, quantifies model discrimination, such that a value of 0.5 indicates that discrimination is no better than random, while a value of 1 would indicate perfect prediction. We used five-fold cross-validation to obtain model performance metrics. Five-fold cross-validation is an internal validation method (*i.e.*, not using an external validation dataset), in which the dataset is split into five groups of equal size. The algorithm is fit on four of the datasets and the fifth dataset is used as the testing dataset to compute performance metrics (James et al., 2013). We repeated training and testing for five random splits of the data and computed the test mean AUC and its standard deviation over the five splits. In other words, we perform all experiments five times (each time leaving a different part of the data for testing) and all performance metrics we report are based on these five runs (we report mean and std).

We also evaluated model performance using the weighted-F1 score. The F1-score is computed as the harmonic mean of positive predictive value and sensitivity, such that the highest value (1.0) indicates both perfect positive predictive value and sensitivity and the lowest value (0) indicates that either the positive predictive value or the sensitivity is zero. To account for imbalance in the data (*i.e.*, differences in the proportions of participants who did and did not conceive), we computed a weighted F1-score as the average of the F1-scores for participants with and without a

pregnancy, weighted by the number of participants in each class. While the AUC is more easily interpretable, the weighted F1-score is more robust to data imbalances (Saito and Rehmsmeier, 2015). Finally, we present weighted-precision and weighted-recall metrics. Precision is equivalent to a positive predictive value, and recall is equivalent to sensitivity. To compute these metrics, we compute precision and recall for each class (*i.e.*, pregnant versus non-pregnant) and calculate their average weighted by the number of true instances for each class.

For Model III, we evaluated performance using the concordance index, as described above.

All analyses were performed with Python statistical functions. Relevant programs can be accessed here: <https://github.com/noc-lab/Predictive-models-of-pregnancy>. Additional methodological information on how we addressed imbalance in the data and tuning of hyperparameters is provided in the Supplementary Text.

## Results

After excluding participants with incomplete follow-up for Models I and II, we analyzed data from 3,195 and 3,476 participants for Models I and II, respectively, and 16,876 cycles from 4,133 participants for Model III. The study participants were aged 30 years on average and ranged in age from 21 to 45 years. Among the 3,195 participants included in Model I, 2,747 (86%) became pregnant in 12 menstrual cycles. Among the 3,476 participants included in Model II, 2,406 (69%) became pregnant within six menstrual cycles. The distribution of class (*i.e.*, pregnant versus non-pregnant), overall and by number of menstrual cycles of attempt time at study entry, are presented in Tables SVI and SVII. For each of the three models, the same

variables were considered for preprocessing (Table I). After statistical feature selection, 40 variables were selected into the full model predicting pregnancy in 12 menstrual cycles (Model I) and 41 variables were selected into the full model predicting pregnancy within six menstrual cycles (Model II). After recursive feature elimination, 30 and 25 variables were selected for the sparse Models I and II, respectively. The final parsimonious models included 14 and 15 variables for Models I and II, respectively. We present performance statistics for the parsimonious models in Table II. The AUC for Model I was 68-70% for all classification algorithms considered (std: 0.8% to 1.9%). The AUCs for Model II were 65-66% (std: 1.9% to 2.6%). The L2LR algorithm generally yielded the highest AUC. The weighted-F1 scores were similar across each algorithm, and no algorithm consistently yielded the highest score. The weighted-F1 scores obtained with the L2LR algorithm were 81.8 (std: 1.0) for Model I and 67.4 (std: 1.4) for Model II. The parsimonious models performed similarly to the full and sparse models (Supplementary Table SI). The concordance index for Model III was 63.5% for the full model after statistical feature selection (24 variables) and 62.6% for the final parsimonious model. Figure SI presents area under the precision-recall curves for Models I, II, IV, and V.

In order of decreasing magnitude of the regression coefficients (*i.e.*, strongest to weakest predictor), the variables selected into the parsimonious Model I that were positively associated with pregnancy were menstrual cycle length, living in a rural region, daily use of multivitamins or folic acid, using the hormonal intrauterine device (IUD) as one's most recent method of contraception, having previously breastfed an infant, having ever been pregnant, female education, recent influenza vaccination, and gravidity (total number of pregnancies) (Table III). The variables that were inversely associated with pregnancy were female age, having a history of

infertility, having completed one menstrual cycle of pregnancy attempt time at study entry (versus zero), female BMI, and stress. The distributions of these variables overall, and by pregnancy status, are presented in Table III. Results for parsimonious Models II and III are presented in Tables IV and V, respectively. The variables selected into the parsimonious Model II that were positively associated with pregnancy were daily use of multivitamins or folic acid, having previously breastfed an infant, HEI-2010 score, having a previous unplanned pregnancy, trying to improve one's chances of pregnancy (*e.g.*, charting cycles, ovulation or cervical mucus testing, timing intercourse to the fertile window), and the time since a participant's last pregnancy (<1 year). The variables that were inversely associated with pregnancy were female BMI, having a history of infertility, male age, non-use of a fertility app, male BMI, having completed one menstrual cycle of pregnancy attempt time at study entry (versus zero), male partner smoking, female age, and having a history of subfertility or infertility. Results were generally similar for Model III. Variables selected into Model III but neither Models I nor II included intercourse frequency, female education, and menstrual cycle regularity.

Among 1,957 nulligravid women without a history of infertility, we developed models predicting pregnancy in fewer than 12 menstrual cycles (Model IV), predicting pregnancy within six menstrual cycles (Model V), and predicting fecundability (Model VI). We analyzed data from 1,571, 1,722, and 1,957 participants for Models IV, V, and VI, respectively. The performance of these models was slightly lower than the analogous models in the full cohort. The performance statistics over five validation subsets for the full and sparse Models IV and V are presented in Table SII. Using statistical feature selection, 16 and 12 variables were selected into the full models for Model IV and V, respectively. After recursive feature elimination, 5 and 9 variables

were selected for the sparse Models IV and V, respectively. Because fewer than 15 features were selected by each of the sparse models, the sparse models were equivalent to the parsimonious models. Consistent with the main analysis, the L2LR algorithm performed best for the sparse models. The AUCs were 69.5% (std: 1.4) for Model IV and 65.6% (std: 2.9) for Model V. The concordance index for Model VI was 60.2%. Variables selected by these models that were positively associated with pregnancy included menstrual cycle length, using a hormonal IUD as one's most recent method of contraception, intercourse frequency, trying to improve one's chances of pregnancy, and HEI-2010 score. Variables inversely associated with the probability of pregnancy included having completed one menstrual cycle of pregnancy attempt time at study entry (versus zero), female age, male and female BMI, menstrual cycle irregularity, non-use of a fertility app, stress, depressive symptoms, history of vaginosis, male partner smoking, milk consumption, and sleep characteristics. Occupational exposures including exposure to metal particulates or fumes and exposure to high temperature environments were also selected to Model VI, but with very small coefficients.

## **Discussion**

In this prospective cohort study of 4,133 pregnancy planners, we applied several supervised learning methods to predict the probability of pregnancy within three time periods: 12 and 6 menstrual cycles, and on a per-cycle basis. The L2LR algorithm generally yielded the highest AUC, particularly for the parsimonious models. Overall, model discrimination (AUC) was close to 70%. These findings demonstrate that it is possible to develop predictive models with reasonable discrimination using self-reported data in the absence of more detailed medical information such as laboratory or imaging tests.

The discrimination of our models is greater than previously published predictive models for pregnancy independent of fertility treatment, which yielded AUC's between 59% and 64% (Coppus, et al., 2009). For example, Eimers *et al.* developed a predictive model for pregnancy among 996 couples consulting for infertility care in the Netherlands between 1974 and 1984 (Eimers et al., 1994). The investigators collected data on patient medical history, laboratory tests including semen analysis and postcoital tests (*i.e.*, an examination of the interaction between sperm and the cervical mucus after intercourse), and a gynecologic physical examination. They used forward stepwise Cox regression to produce a model including female age, duration of infertility, primary versus secondary infertility, history of infertility in the male partner's family, sperm motility, and the postcoital test results. Similar studies were conducted by Collins *et al.*, using data from 1,061 couples seeking infertility care at eleven Canadian University hospitals (Collins et al., 1995), and Snick *et al.*, using data from 402 couples seeking infertility care at a Dutch general hospital (Snick et al., 1997). Hunault *et al.* pooled the data from the Eimers, Collins, and Snick studies to evaluate the accuracy of these models and to develop two new synthesis models (Hunault et al., 2004). The synthesis models included female age, duration of subfertility, sperm motility, whether the couple had been referred for infertility care by a general physician or a gynecologist, and the results of a postcoital test. These models were externally validated and found to have AUCs of 59-63% (Hunault et al., 2005, van der Steeg et al., 2007).

Although previous studies predicted the probability of pregnancy independent of fertility treatment, they were exclusively conducted in populations with subfertility using little or no data on lifestyle, environmental, and sociodemographic factors (Collins, et al., 1995, Coppus, et al.,

2009, Eimers, et al., 1994, Hunault, et al., 2004, Hunault, et al., 2005, Snick, et al., 1997, van der Steeg, et al., 2007). Our study may be more generalizable to couples across the fertility spectrum, because we included couples with a wide range of reproductive potential. In addition, we considered a range of potential predictors that may be more easily modified than clinical markers such as semen quality or hormone levels. For example, fertility app use, use of multivitamins or folic acid supplements, and trying to improve one's chances of pregnancy (*e.g.*, charting cycles, ovulation or cervical mucus testing, timing intercourse to the fertile window) are relatively modifiable behaviors. Lifestyle interventions can also be undertaken to modify individual-level behaviors that may increase a couple's chance of conception, such as promoting a healthy BMI, improving diet, and reducing stress. However, many of these behaviors are determined by environmental and systemic drivers and thus may be best addressed through macro-level policy interventions that address upstream determinants (*e.g.*, regulation of food supply and marketing). A detailed analysis of each factor, from a risk-factor or causal-inference perspective, would be the subject of follow-up and more targeted work. In this study, there were some variables that appeared to be particularly important predictors of pregnancy. These included female age and BMI, history of infertility, the number of menstrual cycles of pregnancy attempt time at study entry, having previously breastfed an infant, and use of multivitamins or folic acid supplements. These findings are generally consistent with previous studies on individual risk factors for infertility that were conducted in other populations (Cueto et al., 2016, Homan, et al., 2007, Jensen et al., 1999, Wise et al., 2011). However, having previously breastfed an infant, which was associated with an increased probability of pregnancy in this study, has not been previously studied as a predictor of fecundability. This may reflect underlying fertility, prolonged effects of

hormonal changes during breastfeeding, or higher socioeconomic status among women who breastfeed their infants (Jones et al., 2011, Odar Stough et al., 2019).

In this study, we developed an additional set of predictive models among nulligravid women with no history of infertility who had been trying to conceive for no more than one menstrual cycle of attempt time at enrollment. The performance of these models was slightly decreased compared with the main analyses. This is likely because having a history of infertility is a strong predictor of future fecundability, and therefore restricting the analytic sample by this variable would limit the predictive ability of the model. This was most obvious in Model V, which predicted pregnancy within six menstrual cycles. In these restricted analyses, the most important predictors of pregnancy across all models were the number of menstrual cycles of pregnancy attempt time at study entry, female age, and stress levels.

Study limitations include potential misclassification of the predictor variables, given that all data were based on self-report. Independent and non-differential misclassification of binary exposure variables is expected, on average, to result in the attenuation of estimates. A 2017 study evaluated the potential influence of misclassification in prognostic prediction models and found that substantial misclassification had a relatively small impact on model accuracy (van Doorn et al., 2017). However, another study reported that measurement error in random forests can result in distorted model performance and variable importance (Jiang et al., 2021). There was also the potential for misspecification of the functional form of the predictor variables, which could have influenced the variable selection process. In addition, there may have been some misclassification of our estimate of time to pregnancy, which relied on self-reported menstrual

cycle length and date of the last menstrual period. Given the prospective design of the study, such misclassification is likely to be non-differential with respect to the outcome. Bias may also have been introduced if the length of follow-up varied by the predictors under study, as Models I and II did not account for varying lengths of follow-up. However, results were generally consistent with Model III, which did account for varying lengths of follow-up. Another potential limitation is that we may not have measured important predictors of pregnancy, such as hormone levels, which could have limited the predictive ability of our models. Other potentially important predictors that we did not measure include environmental exposures (Conforti, et al., 2018, Hipwell et al., 2019, Kahn, et al., 2020), early life adversity (Harville and Boynton-Jarrett, 2013, Jacobs et al., 2015), occupational stress (Barzilai-Pesach et al., 2006, Valsamakis et al., 2019), experiences of discrimination (Krieger, 2000), social disadvantage, neighborhood characteristics (Williams and Collins, 2001), and multigenerational exposures (Eskenazi et al., 2021, Wesselink, 2021). In addition, we lacked comprehensive data on male exposures, which contribute to up to 50% of all subfertility among couples (Irvine, 1998). However, we collected data on several important male characteristics on the female baseline questionnaire, including male age, BMI, education, and smoking status. Overall, we considered a diverse range of 163 potential predictors, which is substantially greater than previous studies in this area (Collins, et al., 1995, Coppus, et al., 2009, Eimers, et al., 1994, Hunault, et al., 2004, Hunault, et al., 2005, Snick, et al., 1997, van der Steeg, et al., 2007). It should be noted that the effect estimates in these models lack causal interpretation, as variables were selected into the final models based on their predictive power, rather than the hypothesized causal structures of the data. Identifying causes of infertility was beyond the scope of this study. Also beyond the scope of this study was the development of models within clinically relevant subgroups (e.g., age or infertility-related

conditions). Finally, though we validated the models using split sample replication techniques, we were unable to conduct an external validation study.

## **Conclusions**

In this large prospective cohort, we used rigorous analytics including machine learning algorithms to develop predictive models of pregnancy, using three distinct, clinically relevant definitions of infertility, subfertility, and fecundability. Comparing results across the three outcomes facilitates robust triangulation of fertility potential; the relative utility of each outcome may depend on a couple's preferences and risk profile. Our methods can predict pregnancy with discrimination as high as 71.2% by properly weighing a small set of predictive variables that include lifestyle and reproductive characteristics. Overall, the most consistent predictors of the probability of conception were female age and BMI, male age and BMI, history of infertility, history of breastfeeding, the time since a participant's last pregnancy, daily use of multivitamins or folic acid, trying to improve one's chances of pregnancy (*e.g.*, charting cycles, ovulation or cervical mucus testing, timing intercourse to the fertile window), male partner smoking, and female education. Among nulligravid women without a history of infertility, the most important predictors were female age and BMI, male BMI, menstrual cycle irregularity, use of a fertility app, and perceived stress. These findings are particularly relevant for couples planning a pregnancy and clinicians providing preconception care to women who are discontinuing contraception in order to conceive. If these models are successfully validated in external populations, they could potentially be implemented as a counseling tool.

**Authors' Roles:** JJY was responsible for results interpretation, manuscript writing, revision, and finalization. TW was responsible for formulation of the study hypotheses and study design, statistical analyses, results interpretation, manuscript writing, revision, and finalization. ZZ was responsible for statistical analyses, results interpretation, manuscript revision, and finalization. SKW was responsible for formulation of the study hypotheses, results interpretation, manuscript revision, and finalization. TRW was responsible for data management, manuscript revision, and finalization. AKW was responsible for formulation of the study hypotheses, results interpretation, manuscript revision, and finalization. TJ was responsible for results interpretation, manuscript revision, and finalization. EEH was responsible for formulation of the study hypotheses and study design, results interpretation, manuscript writing, revision, and finalization. LAW was responsible for formulation of the study hypotheses and study design, statistical oversight, results interpretation, manuscript writing, revision, and finalization. ICP was responsible for formulation of the study hypotheses and study design, statistical oversight, results interpretation, manuscript writing, revision, and finalization.

**Acknowledgments:** We acknowledge the contributions of PRESTO participants and staff.

**Funding:** The research was partially supported by the NSF under grants IIS-1914792, DMS-1664644, and CNS-1645681 and by the NIH under grants R01 GM135930, R21 HD072326, R01 HD086742 and grant 1UL1TR001430 to the Clinical & Translational Science Institute at Boston University.

**Conflict of Interest:** LAW has received in-kind donations from Sandstone Diagnostics, Swiss Precision Diagnostics, FertilityFriend.com and Kindara.com for primary data collection in PRESTO. LAW serves as a fibroid consultant for AbbVie Inc. The other authors have nothing to disclose.

## References

- Akhter S, Marcus M, Kerber RA, Kong M, Taylor KC. The impact of periconceptional maternal stress on fecundability. *Annals of epidemiology* 2016;26: 710-716.e717.
- Barzilai-Pesach V, Sheiner EK, Sheiner E, Potashnik G, Shoham-Vardi I. The effect of women's occupational psychologic stress on outcome of fertility treatments. *Journal of occupational and environmental medicine* 2006;48: 56-62.
- Best D, Bhattacharya S. Obesity and fertility. *Hormone molecular biology and clinical investigation* 2015;24: 5-10.
- Brisimi TS, Xu T, Wang T, Dai W, Adams WG, Paschalidis IC. Predicting chronic disease hospitalizations from electronic health records: An interpretable classification approach. *Proc IEEE Inst Electr Electron Eng* 2018;106: 690-707.
- Buck Louis GM, Barr DB, Kannan K, Chen Z, Kim S, Sundaram R. Paternal exposures to environmental chemicals and time-to-pregnancy: Overview of results from the life study. *Andrology* 2016;4: 639-647.
- Chandra A, Copen CE, Stephen EH. Infertility and impaired fecundity in the united states, 1982-2010: Data from the national survey of family growth. *National health statistics reports* 2013: 1-18, 11 p following 19.
- Chen R, Paschalidis IC. A robust learning approach for regression models based on distributionally robust optimization. *Journal of Machine Learning Research* 2018;19.
- Chen R, Paschalidis IC, Hatabu H, Valtchinov VI, Siegelman J. Detection of unwarranted ct radiation exposure from patient and imaging protocol meta-data using regularized regression. *European journal of radiology open* 2019;6: 206-211.
- Cochran WG. The chi2 test of goodness of fit. *The Annals of Mathematical Statistics* 1952;23: 315-345.
- Collins JA, Burrows EA, Wilan AR. The prognosis for live birth among untreated infertile couples. *Fertil Steril* 1995;64: 22-28.

510 Conforti A, Mascia M, Cioffi G, De Angelis C, Coppola G, De Rosa P, Pivonello R, Alviggi C, De  
 511 Placido G. Air pollution and female fertility: A systematic review of literature. *Reproductive biology and*  
 512 *endocrinology* : *RB&E* 2018;16: 117.  
 513 Coppus SF, van der Veen F, Opmeer BC, Mol BW, Bossuyt PM. Evaluating prediction models in  
 514 reproductive medicine. *Human reproduction (Oxford, England)* 2009;24: 1774-1778.  
 515 Cortes C, Vapnik V. Support-vector networks. *Machine learning* 1995;20: 273-297.  
 516 Cueto HT, Riis AH, Hatch EE, Wise LA, Rothman KJ, Sørensen HT, Mikkelsen EM. Folic acid  
 517 supplementation and fecundability: A danish prospective cohort study. *European journal of clinical*  
 518 *nutrition* 2016;70: 66-71.  
 519 Eimers JM, te Velde ER, Gerritse R, Vogelzang ET, Looman CW, Habbema JD. The prediction of the  
 520 chance to conceive in subfertile couples. *Fertil Steril* 1994;61: 44-52.  
 521 Eskenazi B, Ames J, Rauch S, Signorini S, Brambilla P, Mocarelli P, Siracusa C, Holland N, Warner M.  
 522 Dioxin exposure associated with fecundability and infertility in mothers and daughters of seveso, italy.  
 523 *Human reproduction (Oxford, England)* 2021;36: 794-807.  
 524 Evans-Hoeker EA, Eisenberg E, Diamond MP, Legro RS, Alvero R, Coutifaris C, Casson PR, Christman  
 525 GM, Hansen KR, Zhang H *et al.* Major depression, antidepressant use, and male and female fertility.  
 526 *Fertil Steril* 2018;109: 879-887.  
 527 Evers JL. Female subfertility. *Lancet (London, England)* 2002;360: 151-159.  
 528 Fan D, Liu L, Xia Q, Wang W, Wu S, Tian G, Liu Y, Ni J, Wu S, Guo X *et al.* Female alcohol  
 529 consumption and fecundability: A systematic review and dose-response meta-analysis. *Scientific reports*  
 530 2017;7: 13815.  
 531 Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate  
 532 descent. *J Stat Softw* 2010;33: 1-22.  
 533 Friedman JH. Stochastic gradient boosting. *Computational statistics & data analysis* 2002;38: 367-378.  
 534 Gaskins AJ, Chavarro JE. Diet and fertility: A review. *Am J Obstet Gynecol* 2018;218: 379-389.

535 Gnoth C, Godehardt E, Frank-Herrmann P, Friol K, Tigges J, Freundl G. Definition and prevalence of  
536 subfertility and infertility. *Human Reproduction* 2005;20: 1144-1147.

537 Guenther PM, Casavale KO, Reedy J, Kirkpatrick SI, Hiza HAB, Kuczynski KJ, Kahle LL, Krebs-Smith  
538 SM. Update of the healthy eating index: Hei-2010. *J Acad Nutr Diet* 2013;113: 569-580.

539 Hao B, Sotudian S, Wang T, Xu T, Hu Y, Gaitanidis A, Breen K, Velmahos GC, Paschalidis IC. Early  
540 prediction of level-of-care requirements in patients with covid-19. *Elife* 2020;9.

541 Harville EW, Boynton-Jarrett R. Childhood social hardships and fertility: A prospective cohort study.  
542 *Annals of epidemiology* 2013;23: 784-790.

543 Hastie T, Tibshirani R, Friedman J. *The elements of statistical learning: Data mining, inference, and*  
544 *prediction*, 2009. Springer Science & Business Media.

545 Hawkins DM. The problem of overfitting. *Journal of Chemical Information and Computer Sciences*  
546 2004;44: 1-12.

547 Hipwell AE, Kahn LG, Factor-Litvak P, Porucznik CA, Siegel EL, Fichorova RN, Hamman RF, Klein-  
548 Fedyshin M, Harley KG. Exposure to non-persistent chemicals in consumer products and fecundability: A  
549 systematic review. *Human reproduction update* 2019;25: 51-71.

550 Homan GF, Davies M, Norman R. The impact of lifestyle factors on reproductive performance in the  
551 general population and those undergoing infertility treatment: A review. *Human reproduction update*  
552 2007;13: 209-223.

553 Hunault CC, Habbema JD, Eijkemans MJ, Collins JA, Evers JL, te Velde ER. Two new prediction rules  
554 for spontaneous pregnancy leading to live birth among subfertile couples, based on the synthesis of three  
555 previous models. *Human reproduction (Oxford, England)* 2004;19: 2019-2026.

556 Hunault CC, Laven JS, van Rooij IA, Eijkemans MJ, te Velde ER, Habbema JD. Prospective validation  
557 of two models predicting pregnancy leading to live birth among untreated subfertile couples. *Human*  
558 *reproduction (Oxford, England)* 2005;20: 1636-1641.

559 Irvine DS. Epidemiology and aetiology of male infertility. *Human reproduction* 1998;13: 33-44.

560 Jacobs MB, Boynton-Jarrett RD, Harville EW. Adverse childhood event experiences, fertility difficulties  
 561 and menstrual cycle characteristics. *Journal of psychosomatic obstetrics and gynaecology* 2015;36: 46-  
 562 57.

563 James G, Witten D, Hastie T, Tibshirani R. *An introduction to statistical learning: With applications in r.*  
 564 1 edn, 2013. Springer-Verlag New York.

565 Jensen TK, Scheike T, Keiding N, Schaumburg I, Grandjean P. Fecundability in relation to body mass  
 566 and menstrual cycle patterns. *Epidemiology (Cambridge, Mass)* 1999;10: 422-428.

567 Jiang T, Gradus JL, Lash TL, Fox MP. Addressing measurement error in random forests using  
 568 quantitative bias analysis. *American journal of epidemiology* 2021.

569 Jiang T, Gradus JL, Rosellini AJ. Supervised machine learning: A brief primer. *Behavior Therapy*  
 570 2020;51: 675-687.

571 Joffe M, Key J, Best N, Keiding N, Scheike T, Jensen TK. Studying time to pregnancy by use of a  
 572 retrospective design. *American journal of epidemiology* 2005;162: 115-124.

573 Jones JR, Kogan MD, Singh GK, Dee DL, Grummer-Strawn LM. Factors associated with exclusive  
 574 breastfeeding in the united states. *Pediatrics* 2011;128: 1117-1125.

575 Kahn LG, Harley KG, Siegel EL, Zhu Y, Factor-Litvak P, Porucznik CA, Klein-Fedyshin M, Hipwell  
 576 AE. Persistent organic pollutants and couple fecundability: A systematic review. *Human reproduction*  
 577 *update* 2020.

578 Kingma DP, Ba J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* 2014.

579 Krieger N. Discrimination and health *Social epidemiology*. 2000. Oxford University Press, pp. 36-75.

580 Lee S-I, Lee H, Abbeel P, Ng AY. Efficient l1 regularized logistic regression *Aaai*. 2006, pp. 401-408.

581 Longato E, Vettoretti M, Di Camillo B. A practical perspective on the concordance index for the  
 582 evaluation and selection of prognostic time-to-event models. *Journal of Biomedical Informatics*  
 583 2020;108: 103496.

584 Louis GM, Lum KJ, Sundaram R, Chen Z, Kim S, Lynch CD, Schisterman EF, Pyper C. Stress reduces  
 585 conception probabilities across the fertile window: Evidence in support of relaxation. *Fertil Steril*  
 586 2011;95: 2184-2189.

587 Lynch CD, Sundaram R, Maisog JM, Sweeney AM, Buck Louis GM. Preconception stress increases the  
 588 risk of infertility: Results from a couple-based prospective cohort study--the life study. *Human*  
 589 *reproduction (Oxford, England)* 2014;29: 1067-1075.

590 Macaluso M, Wright-Schnapp TJ, Chandra A, Johnson R, Satterwhite CL, Pulver A, Berman SM, Wang  
 591 RY, Farr SL, Pollack LA. A public health focus on infertility prevention, detection, and management.  
 592 *Fertil Steril* 2010;93: 16.e11-10.

593 Mason L, Baxter J, Bartlett P, Frean M. Boosting algorithms as gradient descent in function space. 1999.  
 594 Nips.

595 Massey FJ. The kolmogorov-smirnov test for goodness of fit. *Journal of the American Statistical*  
 596 *Association* 1951;46: 68-78.

597 Millen AE, Midthune D, Thompson FE, Kipnis V, Subar AF. The national cancer institute diet history  
 598 questionnaire: Validation of pyramid food servings. *American journal of epidemiology* 2006;163: 279-  
 599 288.

600 Nillni YI, Wesselink AK, Gradus JL, Hatch EE, Rothman KJ, Mikkelsen EM, Wise LA. Depression,  
 601 anxiety, and psychotropic medication use and fecundability. *Am J Obstet Gynecol* 2016;215: 453.e451-  
 602 458.

603 Odar Stough C, Khalsa AS, Nabors LA, Merianos AL, Peugh J. Predictors of exclusive breastfeeding for  
 604 6 months in a national sample of us children. *American journal of health promotion : AJHP* 2019;33: 48-  
 605 56.

606 Ripley BD. *Pattern recognition and neural networks*, 2007. Cambridge university press.

607 Saito T, Rehmsmeier M. The precision-recall plot is more informative than the roc plot when evaluating  
 608 binary classifiers on imbalanced datasets. *PloS one* 2015;10: e0118432.

609 Salcedo-Sanz S. Modern meta-heuristics based on nonlinear physics processes: A review of models and  
610 design procedures. *Physics Reports* 2016;655: 1-70.

611 Schmid M, Wright MN, Ziegler A. On the use of harrell's c for clinical risk prediction via random  
612 survival forests. *Expert Systems with Applications* 2016;63: 450-459.

613 Snick HK, Snick TS, Evers JL, Collins JA. The spontaneous pregnancy prognosis in untreated subfertile  
614 couples: The walcheren primary care study. *Human reproduction (Oxford, England)* 1997;12: 1582-1588.

615 Snijder CA, te Velde E, Roeleveld N, Burdorf A. Occupational exposure to chemical substances and time  
616 to pregnancy: A systematic review. *Human reproduction update* 2012;18: 284-300.

617 Soares SR, Melo MA. Cigarette smoking and reproductive function. *Current opinion in obstetrics &*  
618 *gynecology* 2008;20: 281-291.

619 Subar AF, Thompson FE, Kipnis V, Midthune D, Hurwitz P, McNutt S, McIntosh A, Rosenfeld S.  
620 Comparative validation of the block, willett, and national cancer institute food frequency questionnaires :  
621 The eating at america's table study. *American journal of epidemiology* 2001;154: 1089-1099.

622 Sundaram R, Mumford SL, Buck Louis GM. Couples' body composition and time-to-pregnancy. *Human*  
623 *reproduction (Oxford, England)* 2017;32: 662-668.

624 Thoma ME, McLain AC, Louis JF, King RB, Trumble AC, Sundaram R, Louis GMB. Prevalence of  
625 infertility in the united states as estimated by the current duration approach and a traditional constructed  
626 approach. *Fertility and sterility* 2013;99: 1324-1331. e1321.

627 Valsamakis G, Chrousos G, Mastorakos G. Stress, female reproduction and pregnancy.  
628 *Psychoneuroendocrinology* 2019;100: 48-57.

629 van der Steeg JW, Steures P, Eijkemans MJ, Habbema JD, Hompes PG, Broekmans FJ, van Dessel HJ,  
630 Bossuyt PM, van der Veen F, Mol BW. Pregnancy is predictable: A large-scale prospective external  
631 validation of the prediction of spontaneous pregnancy in subfertile couples. *Human reproduction (Oxford,*  
632 *England)* 2007;22: 536-542.

633 van Doorn S, Brakenhoff TB, Moons KGM, Rutten FH, Hoes AW, Groenwold RHH, Geersing GJ. The  
 634 effects of misclassification in routine healthcare databases on the accuracy of prognostic prediction  
 635 models: A case study of the cha2ds2-vasc score in atrial fibrillation. *Diagnostic and Prognostic Research*  
 636 2017;1: 18.

637 Wang T, Paschalidis A, Liu Q, Liu Y, Yuan Y, Paschalidis IC. Predictive models of mortality for  
 638 hospitalized patients with covid-19: Retrospective cohort study. *JMIR Med Inform* 2020;8: e21788.

639 Wesselink AK. Multigenerational effects of environmental exposures. *Human reproduction (Oxford,*  
 640 *England)* 2021;36: 539-542.

641 Wesselink AK, Hatch EE, Rothman KJ, Mikkelsen EM, Aschengrau A, Wise LA. Prospective study of  
 642 cigarette smoking and fecundability. *Human reproduction (Oxford, England)* 2019;34: 558-567.

643 Wesselink AK, Hatch EE, Rothman KJ, Weuve JL, Aschengrau A, Song RJ, Wise LA. Perceived stress  
 644 and fecundability: A preconception cohort study of north american couples. *American journal of*  
 645 *epidemiology* 2018;187: 2662-2671.

646 Wesselink AK, Rothman KJ, Hatch EE, Mikkelsen EM, Sørensen HT, Wise LA. Age and fecundability in  
 647 a north american preconception cohort study. *American journal of obstetrics and gynecology* 2017;217:  
 648 667. e661-667. e668.

649 Williams DR, Collins C. Racial residential segregation: A fundamental cause of racial disparities in  
 650 health. *Public health reports (Washington, DC : 1974)* 2001;116: 404-416.

651 Willis SK, Hatch EE, Wesselink AK, Rothman KJ, Mikkelsen EM, Wise LA. Female sleep patterns, shift  
 652 work, and fecundability in a north american preconception cohort study. *Fertil Steril* 2019;111: 1201-  
 653 1210.e1201.

654 Wise LA, Mikkelsen EM, Rothman KJ, Riis AH, Sørensen HT, Huybrechts KF, Hatch EE. A prospective  
 655 cohort study of menstrual characteristics and time to pregnancy. *American journal of epidemiology*  
 656 2011;174: 701-709.

657 Wise LA, Rothman KJ, Mikkelsen EM, Stanford JB, Wesselink AK, McKinnon C, Gruschow SM,  
 658 Horgan CE, Wiley AS, Hahn KA. Design and conduct of an internet-based preconception cohort study in  
 659 north america: Pregnancy study online. *Paediatric and perinatal epidemiology* 2015;29: 360-371.  
 660 Wise LA, Rothman KJ, Wesselink AK, Mikkelsen EM, Sorensen HT, McKinnon CJ, Hatch EE. Male  
 661 sleep duration and fecundability in a north american preconception cohort study. *Fertil Steril* 2018;109:  
 662 453-459.  
 663 Wise LA, Wesselink AK, Hatch EE, Weuve J, Murray EJ, Wang TR, Mikkelsen EM, Toft Sørensen H,  
 664 Rothman KJ. Changes in behavior with increasing pregnancy attempt time: A prospective cohort study.  
 665 *Epidemiology (Cambridge, Mass)* 2020;31: 659-667.  
 666 Yland JJ, Bresnick KA, Hatch EE, Wesselink AK, Mikkelsen EM, Rothman KJ, Sørensen HT,  
 667 Huybrechts KF, Wise LA. Pregravid contraceptive use and fecundability: Prospective cohort study. *BMJ*  
 668 *(Clinical research ed)* 2020;371: m3966.  
 669

**Table I. Complete list of variables included in analysis.**

| Category                                                                       | Variables Included in Preliminary Analysis                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|--------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Demographic and socioeconomic characteristics                                  | Age, marital status, race, <sup>1</sup> ethnicity, region of residence, urbanization of residential area, year at study entry, highest level of education, parents' education level, household income, employment status, hours/week of work, shift work, night shift frequency in the past month.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| Lifestyle, behavioral, and wellness factors                                    | Cigarette smoking (if so, number per day); total duration of smoking; history of smoking during pregnancy; use of e-cigarettes (if so, ml/day); frequency of marijuana use; exposure to second-hand smoke; alcohol intake; caffeine consumption; moderate physical activity; vigorous physical activity; sedentary activity; sleep duration; trouble sleeping; perceived stress scale score; major depression inventory score.                                                                                                                                                                                                                                                                                                                                                                                                            |
| Dietary factors and use of supplements                                         | Healthy Eating Index-2010 score; supplemental intake of vitamins A, B1, B2, B3, B4, B5, B6, B7, B12, C, D, E, K; beta-carotene; folic acid; iron; zinc; calcium; magnesium; selenium; omega-3 fatty acids; consumption of whole milk, 2% milk, 1% milk, skim milk, soy milk, other milk, fruit juice, bottled water, tap water, sugar-sweetened soda, diet soda, sugar-sweetened energy drinks, diet energy drinks; use of multivitamins or folic acid supplements.                                                                                                                                                                                                                                                                                                                                                                       |
| Early life exposures and family history                                        | Adopted; number of siblings; multiple gestation; born preterm; born with low birthweight; breastfed; delivered via cesarean section; mother's cigarette smoking during pregnancy; mother's age at participant's birth; mother's history of pregnancy complications, miscarriage.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| Reproductive characteristics and disorders                                     | Age at menarche; menstrual regularity; menstrual period characteristics (typical length, number of flow days, flow amount, pain); received human papillomavirus vaccine; abnormal pap smear; ever diagnosed with a thyroid condition, fibroids, polycystic ovarian syndrome, endometriosis, a urinary tract infection, pelvic inflammatory disease, chlamydia, herpes, vaginosis, genital warts; recent use of medications for polycystic ovarian syndrome; gravidity; parity; history of cesarean section; years since last pregnancy; history of unplanned pregnancy; history of subfertility or infertility; history of infertility treatment; history of breastfeeding; number of lifetime sexual partners; doing something to improve pregnancy chances; intercourse frequency; using a fertility app; last method of contraception. |
| Physical characteristics, non-reproductive medical history, and medication use | Body mass index; waist measure; Ferriman-Gallwey Hirsutism Score; handedness; number of primary care visits last year; high blood pressure; received influenza vaccine last year; ever diagnosed with migraines (if so, recent migraine frequency), asthma, hay fever, depression, anxiety, gastroesophageal reflux disease, diabetes; use of the following medications in the 4 weeks before baseline: pain medications, antibiotics, asthma medications, diabetes medications; use of psychotropic medications.                                                                                                                                                                                                                                                                                                                         |
| Environmental exposures (occupational and personal care product use)           | Exposed regularly to agricultural pesticides; metal particulates or fumes; solvents, oil-based paints, or cleaning compounds; high temperature environments; chemotherapeutic drugs; engine exhaust; chemicals for hair dyeing, straightening, or curing; chemicals for manicure/pedicure. Use of chemical hair relaxer.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| Male partner characteristics                                                   | Age, body mass index, education, cigarette smoking (if so, number per day), circumcision status.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |

<sup>1</sup> We conceptualized race as a social construct that serves as a rough proxy for exposure to interpersonal and structural racism.

**Table II. Performance metrics for the parsimonious models, PRESTO 2013-2019.**

| Algorithm <sup>1</sup> | Performance Measure (%) (Standard Deviation) |                   |                    |                 |                 |                   |                    |                 |
|------------------------|----------------------------------------------|-------------------|--------------------|-----------------|-----------------|-------------------|--------------------|-----------------|
|                        | <i>Model I</i>                               |                   |                    |                 | <i>Model II</i> |                   |                    |                 |
|                        | AUC                                          | Weighted F1 Score | Weighted Precision | Weighted Recall | AUC             | Weighted F1 Score | Weighted Precision | Weighted Recall |
| L2LR                   | 70.2 (1.6)                                   | 81.8 (1.0)        | 80.8 (1.0)         | 83.3 (1.3)      | 66.1 (2.1)      | 67.5 (1.6)        | 67.2 (1.5)         | 69.5 (1.4)      |
| L1LR                   | 69.8 (1.8)                                   | 81.6 (0.6)        | 80.6 (0.8)         | 83.5 (1.1)      | 66.0 (1.9)      | 67.4 (1.6)        | 67.0 (1.6)         | 69.3 (1.5)      |
| L1SVM                  | 69.8 (1.9)                                   | 81.8 (0.8)        | 80.6 (0.8)         | 83.6 (0.8)      | 66.0 (1.9)      | 67.4 (1.2)        | 66.9 (1.3)         | 69.1 (1.3)      |
| L2SVM                  | 70.0 (1.6)                                   | 81.5 (1.1)        | 80.5 (1.2)         | 83.3 (1.3)      | 66.2 (2.1)      | 67.2 (1.0)        | 66.9 (1.1)         | 69.6 (0.9)      |
| MLP                    | 69.9 (0.8)                                   | 82.1 (0.9)        | 81.1 (1.2)         | 83.9 (1.3)      | 65.1 (2.1)      | 67.5 (1.5)        | 67.0 (1.5)         | 68.5 (1.7)      |
| LightGBM               | 68.1 (1.4)                                   | 81.6 (0.8)        | 80.8 (0.9)         | 82.9 (1.2)      | 64.9 (2.6)      | 66.9 (1.3)        | 66.6 (1.4)         | 67.6 (1.1)      |

Note: Model I predicts pregnancy in <12 menstrual cycles (N = 3,195 participants). Model II predicts pregnancy in <7 menstrual cycles (N = 3,476 participants). The parsimonious models contain variables selected after both statistical feature selection and recursive feature elimination, and limiting recursive feature elimination to select a model with up to 15 variables.

<sup>1</sup> L2LR,  $\ell_2$ -penalized logistic regression; L1LR,  $\ell_1$ -penalized logistic regression; L1SVM, Support Vector Machine (SVM) with an  $\ell_1$ -norm regularizer; L2SVM, SVM with an  $\ell_2$ -norm regularizer; MLP, Feed Forward Multilayer Perceptron Neural Networks; LightGBM, Light Gradient Boosting Machine.

**Table III. Variables selected by the parsimonious Model I (predicting pregnancy in 12 cycles) using the L2LR algorithm, PRESTO 2013-2019, n=3,195 participants.**

| Variable                                                                                     | Standardized<br>Regression<br>Coefficient | Overall              |      | Pregnant             |      | Not pregnant         |      |
|----------------------------------------------------------------------------------------------|-------------------------------------------|----------------------|------|----------------------|------|----------------------|------|
|                                                                                              |                                           | Frequency<br>or Mean | Std. | Frequency<br>or Mean | Std. | Frequency<br>or Mean | Std. |
| Menstrual cycle length (days)                                                                | 0.27                                      | 29.6                 | 4.0  | 29.7                 | 4.1  | 28.7                 | 3.0  |
| Female age at baseline (years)                                                               | -0.26                                     | 29.8                 | 3.8  | 29.7                 | 3.6  | 30.6                 | 4.5  |
| Urbanization of residential area: rural (ref = urbanized area)                               | 0.25                                      | 4%                   | 20%  | 5%                   | 21%  | 1%                   | 12%  |
| Previously tried to conceive for $\geq 12$ months: “yes” (ref = “no, tried for < 12 months”) | -0.24                                     | 5%                   | 21%  | 4%                   | 19%  | 10%                  | 30%  |
| One menstrual cycle of attempt time at study entry (ref = 0)                                 | -0.23                                     | 58%                  | 49%  | 56%                  | 50%  | 68%                  | 47%  |
| Daily use of multivitamins/folic acid (yes/no)                                               | 0.22                                      | 84%                  | 37%  | 85%                  | 35%  | 73%                  | 44%  |
| Last method of contraception: hormonal IUD (yes/no) <sup>1</sup>                             | 0.19                                      | 12%                  | 32%  | 12%                  | 33%  | 7%                   | 25%  |
| Female BMI (kg/m <sup>2</sup> )                                                              | -0.19                                     | 26.6                 | 6.5  | 26.3                 | 6.2  | 28.4                 | 7.8  |
| Ever breastfed an infant (yes/no)                                                            | 0.18                                      | 31%                  | 46%  | 32%                  | 47%  | 22%                  | 41%  |
| Ever been pregnant (yes/no)                                                                  | 0.15                                      | 50%                  | 50%  | 52%                  | 50%  | 42%                  | 49%  |
| Female education (years)                                                                     | 0.14                                      | 16.0                 | 1.2  | 16.1                 | 1.2  | 15.8                 | 1.4  |
| Received influenza vaccine in the past year (yes/no)                                         | 0.13                                      | 53%                  | 50%  | 54%                  | 50%  | 44%                  | 50%  |
| Stress (Perceived Stress Scale score)                                                        | -0.12                                     | 15.5                 | 5.8  | 15.3                 | 5.8  | 16.3                 | 5.6  |
| Total number of pregnancies                                                                  | 0.12                                      | 1.0                  | 1.4  | 1.0                  | 1.4  | 0.8                  | 1.4  |
| <b>Variables forced into the model<sup>2</sup></b>                                           |                                           |                      |      |                      |      |                      |      |
| Urbanization of residential area: Canada (ref = urbanized area)                              | 0.01                                      | 18%                  | 39%  | 18%                  | 39%  | 19%                  | 39%  |
| Urbanization of residential area: urban cluster (ref = urbanized area)                       | -0.01                                     | 8%                   | 27%  | 8%                   | 27%  | 8%                   | 27%  |

|                                                                                                                   |       |     |     |     |     |     |     |
|-------------------------------------------------------------------------------------------------------------------|-------|-----|-----|-----|-----|-----|-----|
| Previously tried to conceive for $\geq 12$ months: “no, never tried before” (ref = “no, tried for $< 12$ months”) | -0.01 | 42% | 49% | 41% | 49% | 48% | 50% |
|-------------------------------------------------------------------------------------------------------------------|-------|-----|-----|-----|-----|-----|-----|

Note: Variables are presented in order of the magnitude of the standardized regression coefficients.

<sup>1</sup> Last methods of contraception were not mutually exclusive and were coded as indicator variables with no reference category. Natural methods included withdrawal, avoiding sex when fertile, calendar methods, and monitoring cervical mucus or basal body temperature.

<sup>2</sup> For all models, we selected a reference group for each categorical variable that was recoded as indicator variables in the preprocessing phase and forced every non-reference level to be included in the model if any level of the categorical variable was selected. These variables are listed in addition to the variables selected by the parsimonious model.

**Table IV. Variables selected by the parsimonious Model II (predicting pregnancy within 6 cycles) using the L2LR algorithm, PRESTO 2013-2019, n=3,476 participants.**

[illegible]

|                                                                                                                 |       |     |     |     |     |     |     |
|-----------------------------------------------------------------------------------------------------------------|-------|-----|-----|-----|-----|-----|-----|
| Previously tried to conceive for $\geq 12$ months: “no, never tried before” (ref = “no, tried for < 12 months”) | -0.05 | 42% | 49% | 40% | 49% | 46% | 50% |
| Time since last pregnancy: 1-2 years (ref = nulliparous)                                                        | 0.04  | 17% | 38% | 19% | 39% | 14% | 35% |
| Male cigarette smoking: “yes, occasionally” (ref = “no”)                                                        | -0.02 | 4%  | 20% | 4%  | 19% | 5%  | 22% |
| Time since last pregnancy: $\geq 5$ years (ref = nulliparous)                                                   | -0.02 | 6%  | 24% | 5%  | 22% | 8%  | 27% |
| Use of fertility app: “no” (ref = “yes”)                                                                        | -0.02 | 23% | 42% | 22% | 41% | 26% | 44% |
| Time since last pregnancy: 3-4 years (ref = nulliparous)                                                        | 0.02  | 4%  | 21% | 5%  | 21% | 4%  | 19% |

Note: Variables are presented in order of the magnitude of the standardized regression coefficients.

<sup>1</sup> For all models, we selected a reference group for each categorical variable that was recoded as indicator variables in the preprocessing phase and forced every non-reference level to be included in the model if any level of the categorical variable was selected. These variables are listed in addition to the variables selected by the parsimonious model.

**Table V. Variables selected by the parsimonious Model III (fecundability), PRESTO 2013-2019, n=4,133 participants.**

| Variable                                                                                                                                            | Hazard Ratio | 95% Confidence Interval |
|-----------------------------------------------------------------------------------------------------------------------------------------------------|--------------|-------------------------|
| Previously tried to conceive for $\geq 12$ months: “yes” (ref = “no, tried for < 12 months”)                                                        | 0.85         | (0.80, 0.90)            |
| Ever breastfed an infant (yes/no)                                                                                                                   | 1.16         | (1.09, 1.23)            |
| Female BMI (kg/m <sup>2</sup> )                                                                                                                     | 0.89         | (0.84, 0.93)            |
| Time since last pregnancy: 1-2 years (ref = nulliparous)                                                                                            | 1.12         | (1.04, 1.21)            |
| Female age at baseline (years)                                                                                                                      | 0.90         | (0.85, 0.95)            |
| Trying to improve chances of pregnancy (yes/no)                                                                                                     | 1.11         | (1.06, 1.15)            |
| Female education (years)                                                                                                                            | 1.09         | (1.03, 1.15)            |
| Intercourse frequency (times/week)                                                                                                                  | 1.08         | (1.03, 1.12)            |
| Male BMI (kg/m <sup>2</sup> )                                                                                                                       | 0.93         | (0.89, 0.98)            |
| Male cigarette smoking: “yes, on a regular basis” (ref = “no”)                                                                                      | 0.93         | (0.89, 0.98)            |
| Has menstrual cycle been regular without hormonal contraception in past 2 years? “no, irregular” (ref = “yes, regular”)                             | 0.94         | (0.89, 0.99)            |
| Daily use of multivitamins/folic acid (yes/no)                                                                                                      | 1.06         | (1.01, 1.11)            |
| Did your period become regular on its own? “no, irregular” (ref = “yes, regular”)                                                                   | 0.96         | (0.92, 1.01)            |
| Male age (years)                                                                                                                                    | 0.96         | (0.91, 1.01)            |
| Tap water consumption (drinks/week)                                                                                                                 | 1.04         | (1.01, 1.07)            |
| <b>Variables forced into the model<sup>1</sup></b>                                                                                                  |              |                         |
| Time since last pregnancy: <1 year (ref = nulliparous)                                                                                              | 1.37         | (1.14, 1.64)            |
| Time since last pregnancy: 3-4 years (ref = nulliparous)                                                                                            | 1.32         | (1.01, 1.71)            |
| Male cigarette smoking: “yes, occasionally” (ref = “no”)                                                                                            | 0.87         | (0.70, 1.08)            |
| Has menstrual cycle been regular without hormonal contraception in past 2 years? “unknown, was using hormonal contraception” (ref = “yes, regular”) | 1.03         | (0.93, 1.14)            |
| Did your period become regular on its own? “unknown, was using hormonal contraception” (ref = “yes, regular”)                                       | 1.02         | (0.89, 1.17)            |
| Time since last pregnancy: $\geq 5$ years (ref = nulliparous)                                                                                       | 1.01         | (0.79, 1.29)            |

Note: Variables are presented in order of the magnitude of the regression coefficients (*i.e.*, the natural logarithm of the Hazard Ratio)

<sup>1</sup> For all models, we selected a reference group for each categorical variable that was recoded as indicator variables in the preprocessing phase and forced every non-reference level to be included in the model if any level of the categorical variable was selected. These variables are listed in addition to the variables selected by the parsimonious model.

## **Supplementary Material**

Supplement to: Predictive models of pregnancy based on data from a preconception cohort study

Jennifer J Yland; Taiyao Wang; Zahra Zad; Sydney K Willis; Tanran R Wang; Amelia K Wesselink; Tammy Jiang; Elizabeth E Hatch; Lauren A Wise; Ioannis Ch. Paschalidis

## Table of Contents

| Item               | Description                                                                                                                                                                              | Page |
|--------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|
| Supplementary Text | Supplementary information on modeling approaches.                                                                                                                                        | 2    |
| Table SI           | Performance metrics for the full and sparse Models I-II.                                                                                                                                 | 4    |
| Table SII          | Performance metrics for the full and sparse Models IV-V, restricting to nulligravid women with no history of infertility.                                                                | 5    |
| Table SIII         | Variables selected by the parsimonious Model IV (restricted to nulligravid women with no history of infertility) predicting pregnancy within 12 cycles, using the L2LR algorithm.        | 6    |
| Table SIV          | Variables selected by the parsimonious Model V (restricted to nulligravid women with no history of infertility) predicting pregnancy within 6 cycles, using the L2LR algorithm.          | 7    |
| Table SV           | Variables selected by the parsimonious Model VI (restricted to nulligravid women with no history of infertility) predicting pregnancy within 6 cycles, using the L2LR algorithm, PRESTO. | 8    |
| Table SVI          | Distribution of class and number of menstrual cycles of attempt time at study entry.                                                                                                     | 9    |
| Table SVII         | Distribution of each class by the number of menstrual cycles of attempt time at study entry.                                                                                             | 10   |
| Figure SI          | Area under the precision-recall curves (AUPRC) for Models I, II, IV, and V.                                                                                                              | 11   |
| Figure SII         | Visualization of the model coefficients and their 95% confidence intervals.                                                                                                              | 12   |

## Supplementary Text

### *Addressing the distribution and bias of positive and negative cases in the data*

To address the class imbalance in our dataset, we use a class weight (inversely proportional to class size) in the loss function used for training the model. This has the effect of balancing contributions to the loss from both classes. Class weights are used differently depending on the algorithm: for linear models (such as linear SVM or logistic regression), the class weights alter the loss function by weighting the loss of each sample by its class weight. For tree-based algorithms, the class weights are used for reweighting the splitting criterion. However, this rebalancing does not take the weight of samples in each class into account.

### *Tuning of hyperparameters*

We tune hyperparameters through cross-validation. In Logistic Regression (LR) and Support Vector Machine (SVM) models, we consider the inverse of regularization strength as a hyper parameter. We search for the best hyper parameter among [0.001, 0.01, 0.1, 1, 10] and choose the one that leads to the best classifier (with the highest AUC). In the artificial neural network (MLP) models, we have one input layer, a number of hidden layers, and one output layer. We tune the number of hidden layers and the number of neurons in the hidden layers. We try different options: (i) one hidden layer with 32, 64, 128, 256, or 512 neurons, (ii) two hidden layers with 16, 32, 64, 128, 256 neurons in the first hidden layer and 2 neurons in the second hidden layer, (iii) two hidden layers with 8, 16, 32, 64, or 128 neurons in the first hidden layer and 4 neurons in the second hidden layer. In the Gradient Boosting Machine (GBM) models, we used LightGBM which is a fast and high-performance GBM framework that grows trees leaf-wise rather than level-wise and incorporates advanced techniques, such as gradient-based one-

side sampling and exclusive feature bundling to deal with a large number of data instances and features. We tune a number of hyper parameters such as learning rate, maximum number of leaves in one tree, and minimal number of data in one leaf. More details on the range of numbers used for the tuning of hyper parameters of our LightGBM can be found in the scripts at the github (<https://github.com/noc-lab/Predictive-models-of-pregnancy>).

**Table SI. Performance metrics for the full and sparse Models I-II, PRESTO 2013-2019.**

| Algorithm <sup>1</sup>          | Performance Measure (%) (Standard Deviation) |                   |                    |                 |                 |                   |                    |                 |
|---------------------------------|----------------------------------------------|-------------------|--------------------|-----------------|-----------------|-------------------|--------------------|-----------------|
|                                 | <i>Model I</i>                               |                   |                    |                 | <i>Model II</i> |                   |                    |                 |
|                                 | AUC                                          | Weighted F1 Score | Weighted Precision | Weighted Recall | AUC             | Weighted F1 Score | Weighted Precision | Weighted Recall |
| <b>Full Model<sup>2</sup></b>   |                                              |                   |                    |                 |                 |                   |                    |                 |
| L2LR                            | 70.5 (1.2)                                   | 82.0 (0.5)        | 81.3 (0.6)         | 82.8 (0.5)      | 65.9 (1.3)      | 67.1 (1.4)        | 66.7 (1.3)         | 68.4 (2.0)      |
| L1LR                            | 70.3 (1.4)                                   | 82.7 (0.7)        | 81.9 (0.5)         | 84.1 (1.1)      | 65.5 (1.5)      | 66.8 (0.8)        | 66.4 (0.9)         | 68.1 (1.6)      |
| L1SVM                           | 70.5 (1.5)                                   | 82.6 (0.5)        | 81.7 (0.4)         | 83.9 (0.9)      | 65.2 (1.7)      | 66.3 (0.9)        | 65.9 (0.9)         | 67.0 (1.1)      |
| L2SVM                           | 70.8 (1.4)                                   | 82.2 (0.6)        | 81.5 (0.5)         | 83.2 (1.1)      | 65.6 (1.3)      | 66.8 (1.1)        | 66.4 (1.1)         | 67.9 (1.7)      |
| MLP                             | 69.7 (2.1)                                   | 81.8 (1.0)        | 81.2 (0.8)         | 82.8 (1.8)      | 63.4 (1.4)      | 65.5 (1.0)        | 64.9 (1.2)         | 66.9 (0.8)      |
| LightGBM                        | 68.8 (0.6)                                   | 81.3 (0.6)        | 80.3 (0.4)         | 82.8 (1.0)      | 63.8 (1.9)      | 65.8 (0.9)        | 65.3 (0.9)         | 66.8 (1.0)      |
| <b>Sparse Model<sup>3</sup></b> |                                              |                   |                    |                 |                 |                   |                    |                 |
| L2LR                            | 71.2 (1.0)                                   | 81.8 (0.6)        | 81.1 (0.6)         | 82.8 (0.9)      | 67.1 (1.5)      | 67.7 (0.3)        | 67.3 (0.3)         | 69.4 (0.9)      |
| L1LR                            | 70.5 (1.7)                                   | 82.6 (0.7)        | 81.7 (0.8)         | 84.2 (1.0)      | 66.6 (1.3)      | 67.7 (1.0)        | 67.3 (1.0)         | 68.6 (1.4)      |
| L1SVM                           | 70.7 (1.7)                                   | 82.4 (1.2)        | 81.7 (1.0)         | 83.7 (1.9)      | 66.5 (1.2)      | 67.0 (0.7)        | 66.7 (0.8)         | 68.2 (0.9)      |
| L2SVM                           | 71.2 (1.4)                                   | 82.2 (0.7)        | 81.5 (0.6)         | 83.5 (1.3)      | 66.8 (1.4)      | 67.3 (0.5)        | 66.8 (0.5)         | 68.4 (0.8)      |
| MLP                             | 70.8 (2.6)                                   | 82.5 (1.0)        | 81.5 (1.2)         | 84.5 (0.7)      | 65.3 (0.6)      | 67.0 (1.0)        | 66.6 (0.9)         | 67.7 (1.3)      |
| LightGBM                        | 69.3 (1.3)                                   | 80.8 (0.5)        | 79.9 (0.4)         | 82.2 (1.2)      | 65.0 (2.1)      | 66.3 (1.6)        | 65.9 (1.5)         | 66.9 (1.8)      |

Note: Model I predicts pregnancy in <12 menstrual cycles (N = 3,195 participants). Model II predicts pregnancy in <7 menstrual cycles (N = 3,476 participants).

<sup>1</sup> L2LR,  $\ell_2$ -penalized logistic regression; L1LR,  $\ell_1$ -penalized logistic regression; L1SVM, Support Vector Machine (SVM) with an  $\ell_1$ -norm regularizer; L2SVM, SVM with an  $\ell_2$ -norm regularizer; MLP, Feed Forward Multilayer Perceptron Neural Networks; LightGBM, Light Gradient Boosting Machine.

<sup>2</sup> The full models contain all variables selected after statistical feature selection.

<sup>3</sup> The sparse models contain variables selected after both statistical feature selection and recursive feature elimination.

**Table SII. Performance metrics for the full and sparse Models IV-V, restricting to nulligravid women with no history of infertility, PRESTO 2013-2019.**

| Algorithm <sup>1</sup>          | Performance Measure (%) (Standard Deviation) |                   |                    |                 |                |                   |                    |                 |
|---------------------------------|----------------------------------------------|-------------------|--------------------|-----------------|----------------|-------------------|--------------------|-----------------|
|                                 | <i>Model IV</i>                              |                   |                    |                 | <i>Model V</i> |                   |                    |                 |
|                                 | AUC                                          | Weighted F1 Score | Weighted Precision | Weighted Recall | AUC            | Weighted F1 Score | Weighted Precision | Weighted Recall |
| <b>Full Model<sup>2</sup></b>   |                                              |                   |                    |                 |                |                   |                    |                 |
| L2LR                            | 68.3 (4.1)                                   | 79.2 (1.5)        | 78.7 (2.5)         | 81.1 (2.6)      | 65.1 (3.2)     | 64.5 (1.7)        | 65.2 (2.0)         | 67.5 (2.0)      |
| L1LR                            | 68.4 (4.1)                                   | 78.3 (1.7)        | 77.9 (2.5)         | 79.7 (2.5)      | 64.9 (2.7)     | 65.2 (2.4)        | 65.2 (2.6)         | 67.5 (2.0)      |
| L1SVM                           | 68.4 (3.9)                                   | 78.6 (1.7)        | 77.8 (2.5)         | 80.0 (2.0)      | 64.9 (2.7)     | 65.1 (2.3)        | 65.2 (2.5)         | 67.5 (2.0)      |
| L2SVM                           | 68.6 (3.7)                                   | 79.0 (1.1)        | 78.4 (2.4)         | 80.5 (1.9)      | 64.9 (2.8)     | 64.6 (2.4)        | 65.4 (2.7)         | 67.4 (2.8)      |
| MLP                             | 67.8 (2.4)                                   | 79.3 (1.6)        | 78.6 (1.7)         | 80.4 (1.9)      | 63.9 (3.3)     | 63.9 (2.2)        | 63.7 (2.2)         | 65.8 (2.2)      |
| LightGBM                        | 64.3 (3.2)                                   | 78.6 (1.4)        | 77.6 (1.7)         | 80.1 (1.3)      | 63.7 (3.0)     | 63.8 (2.1)        | 63.4 (2.1)         | 64.8 (2.0)      |
| <b>Sparse Model<sup>3</sup></b> |                                              |                   |                    |                 |                |                   |                    |                 |
| L2LR                            | 69.5 (1.4)                                   | 80.5 (1.0)        | 79.8 (0.8)         | 81.8 (1.5)      | 65.6 (2.9)     | 66.2 (2.5)        | 65.9 (2.6)         | 66.8 (2.4)      |
| L1LR                            | 69.3 (1.3)                                   | 80.6 (0.9)        | 79.9 (0.8)         | 81.9 (1.5)      | 64.9 (2.8)     | 64.7 (2.6)        | 64.9 (2.9)         | 67.1 (2.3)      |
| L1SVM                           | 69.4 (1.3)                                   | 80.7 (1.1)        | 80.0 (0.9)         | 81.9 (1.6)      | 64.8 (2.8)     | 64.8 (2.6)        | 65.0 (3.0)         | 67.1 (2.3)      |
| L2SVM                           | 69.5 (1.5)                                   | 80.6 (1.3)        | 79.8 (1.1)         | 81.8 (1.6)      | 65.2 (2.8)     | 65.4 (2.2)        | 65.3 (2.2)         | 67.3 (2.0)      |
| MLP                             | 68.1 (1.7)                                   | 79.6 (1.4)        | 78.6 (1.9)         | 81.1 (1.3)      | 63.9 (2.8)     | 64.2 (2.8)        | 64.0 (3.0)         | 64.8 (2.3)      |
| LightGBM                        | 66.0 (2.4)                                   | 79.3 (0.5)        | 78.4 (0.3)         | 80.7 (1.2)      | 63.5 (3.3)     | 64.3 (2.2)        | 63.9 (2.2)         | 65.6 (1.8)      |

Note: Model IV predicts pregnancy in <12 menstrual cycles among nulligravid women with no history of infertility (N = 1,571 participants). Model V predicts pregnancy in <7 menstrual cycles among nulligravid women with no history of infertility (N = 1,722 participants).

<sup>1</sup> L2LR,  $\ell_2$ -penalized logistic regression; L1LR,  $\ell_1$ -penalized logistic regression; L1SVM, Support Vector Machine (SVM) with an  $\ell_1$ -norm regularizer; L2SVM, SVM with an  $\ell_2$ -norm regularizer; MLP, Feed Forward Multilayer Perceptron Neural Networks; LightGBM, Light Gradient Boosting Machine.

<sup>2</sup> The full models contain all variables selected after statistical feature selection.

<sup>3</sup> The sparse models contain variables selected after both statistical feature selection and recursive feature elimination.

**Table SIII. Variables selected by the parsimonious Model IV (restricted to nulligravid women with no history of infertility) predicting pregnancy within 12 cycles, using the L2LR algorithm, PRESTO 2013-2019, n=1,571 participants.**

| Variable                                                            | Standardized<br>Regression<br>Coefficient | Overall              |      | Pregnant             |      | Not pregnant         |      |
|---------------------------------------------------------------------|-------------------------------------------|----------------------|------|----------------------|------|----------------------|------|
|                                                                     |                                           | Frequency or<br>Mean | Std. | Frequency<br>or Mean | Std. | Frequency<br>or Mean | Std. |
| Menstrual cycle length (days)                                       | 0.27                                      | 29.6                 | 4.0  | 29.8                 | 4.2  | 28.7                 | 2.6  |
| One menstrual cycle of attempt time at study entry<br>(ref=0)       | -0.22                                     | 58%                  | 49%  | 55%                  | 50%  | 70%                  | 46%  |
| Last method of contraception: hormonal IUD<br>(yes/no) <sup>1</sup> | 0.22                                      | 11%                  | 31%  | 12%                  | 32%  | 5%                   | 22%  |
| Stress (Perceived Stress Scale score)                               | -0.20                                     | 15.0                 | 5.5  | 14.7                 | 5.5  | 16.1                 | 5.3  |
| Female age at baseline (years)                                      | -0.20                                     | 29.2                 | 3.4  | 29.1                 | 3.3  | 30.0                 | 3.9  |

Note: Variables are presented in order of the magnitude of the standardized regression coefficients.

<sup>1</sup> Last methods of contraception were not mutually exclusive and were coded as indicator variables with no reference category.

**Table SIV. Variables selected by the parsimonious Model V (restricted to nulligravid women with no history of infertility) predicting pregnancy within 6 cycles, using the L2LR algorithm, PRESTO 2013-2019, n=1,722 participants.**

| Variable                                                                          | Standardized Regression Coefficient | Overall           |      | Pregnant          |      | Not pregnant      |      |
|-----------------------------------------------------------------------------------|-------------------------------------|-------------------|------|-------------------|------|-------------------|------|
|                                                                                   |                                     | Frequency or Mean | Std. | Frequency or Mean | Std. | Frequency or Mean | Std. |
| Female age at baseline (years)                                                    | -0.06                               | 29.2              | 3.4  | 28.9              | 3.2  | 29.8              | 3.8  |
| Use of fertility app: “no” (ref = “yes”)                                          | -0.05                               | 22%               | 42%  | 20%               | 40%  | 27%               | 44%  |
| Male BMI (kg/m <sup>2</sup> )                                                     | -0.04                               | 27.4              | 5.3  | 27.0              | 5.1  | 28.2              | 5.5  |
| One menstrual cycle of attempt time at study entry (ref = 0)                      | -0.04                               | 58%               | 49%  | 55%               | 50%  | 64%               | 48%  |
| Did your period become regular on its own? “no, irregular” (ref = “yes, regular”) | -0.04                               | 19%               | 39%  | 16%               | 37%  | 23%               | 42%  |
| Healthy Eating Index-2010 score (HEI-2010 score)                                  | 0.04                                | 67.2              | 10.8 | 67.8              | 10.5 | 65.8              | 11.3 |
| Female BMI (kg/m <sup>2</sup> )                                                   | -0.04                               | 26.3              | 6.5  | 25.8              | 6.0  | 27.4              | 7.2  |
| Use of fertility app: “no, but I plan to” (ref = “yes”)                           | -0.03                               | 9%                | 28%  | 8%                | 26%  | 11%               | 31%  |
| Stress (Perceived Stress Scale score)                                             | -0.02                               | 15.1              | 5.6  | 14.8              | 5.5  | 15.6              | 5.6  |
| Depressive Symptoms (Major Depression Inventory score)                            | -0.02                               | 9.0               | 7.0  | 8.7               | 6.9  | 9.6               | 7.0  |

**Variables forced into the model<sup>1</sup>**

|                                                                                                               |      |     |     |     |     |     |     |
|---------------------------------------------------------------------------------------------------------------|------|-----|-----|-----|-----|-----|-----|
| Did your period become regular on its own? “unknown, was using hormonal contraception” (ref = “yes, regular”) | 0.01 | 15% | 35% | 16% | 37% | 12% | 33% |
|---------------------------------------------------------------------------------------------------------------|------|-----|-----|-----|-----|-----|-----|

Note: Variables are presented in order of the magnitude of the standardized regression coefficients.

<sup>1</sup> For all models, we selected a reference group for each categorical variable that was recoded as indicator variables in the preprocessing phase and forced every non-reference level to be included in the model if any level of the categorical variable was selected. These variables are listed in addition to the variables selected by the parsimonious model.

**Table SV. Variables selected by the parsimonious Model VI (restricted to nulligravid women with no history of infertility) predicting pregnancy within 6 cycles, using the L2LR algorithm, PRESTO 2013-2019, n=1,957 participants.**

| Variable                                                                                                                                            | Hazard Ratio | 95% Confidence Interval |
|-----------------------------------------------------------------------------------------------------------------------------------------------------|--------------|-------------------------|
| Female age at baseline (years)                                                                                                                      | 0.90         | (0.84, 0.96)            |
| Has menstrual cycle been regular without hormonal contraception in past 2 years? “no, irregular” (ref = “yes, regular”)                             | 0.90         | (0.84, 0.97)            |
| Use of fertility app: “no” (ref = “yes”)                                                                                                            | 0.90         | (0.85, 0.97)            |
| Ever diagnosed with vaginosis (yes/no)                                                                                                              | 0.91         | (0.85, 0.97)            |
| Intercourse frequency (times/week)                                                                                                                  | 1.10         | (1.03, 1.17)            |
| Male cigarette smoking: “yes, on a regular basis” (ref = “no”)                                                                                      | 0.91         | (0.85, 0.97)            |
| Female BMI (kg/m <sup>2</sup> )                                                                                                                     | 0.91         | (0.85, 0.98)            |
| Exposed regularly to metal particulates or fumes (yes/no)                                                                                           | 1.08         | (1.02, 1.15)            |
| Male BMI (kg/m <sup>2</sup> )                                                                                                                       | 0.94         | (0.88, 1.01)            |
| Trying to improve chances of pregnancy (yes/no)                                                                                                     | 1.06         | (0.99, 1.13)            |
| 2% milk consumption (drinks/week)                                                                                                                   | 0.95         | (0.89, 1.01)            |
| Use of vitamin E supplements (yes/no)                                                                                                               | 1.05         | (0.99, 1.12)            |
| Nightly sleep duration (hours)                                                                                                                      | 0.96         | (0.90, 1.02)            |
| Night shift work (number of shifts in past month)                                                                                                   | 0.96         | (0.90, 1.03)            |
| Has menstrual cycle been regular without hormonal contraception in past 2 years? “unknown, was using hormonal contraception” (ref = “yes, regular”) | 0.98         | (0.86, 1.12)            |
| Exposed regularly to environments with high temperature (yes/no)                                                                                    | 1.02         | (0.95, 1.08)            |
| <b>Variables forced into the model<sup>1</sup></b>                                                                                                  |              |                         |
| Use of fertility app: “no, but I plan to” (ref = “yes”)                                                                                             | 0.76         | (0.60, 0.96)            |
| Male cigarette smoking: “yes, occasionally” (ref = “no”)                                                                                            | 0.86         | (0.62, 1.19)            |
| Has menstrual cycle been regular without hormonal contraception in past 2 years? “unknown, was using hormonal contraception” (ref = “yes, regular”) | 0.98         | (0.86, 1.12)            |

Note: Variables are presented in order of the magnitude of the regression coefficients (*i.e.*, the natural logarithm of the Hazard Ratio)

<sup>1</sup> For all models, we selected a reference group for each categorical variable that was recoded as indicator variables in the preprocessing phase and forced every non-reference level to be included in the model if any level of the categorical variable was selected. These variables are listed in addition to the variables selected by the parsimonious model.

**Table SVI. Distribution of class and number of menstrual cycles of attempt time at study entry.**

|           | N    | Pregnant   | Non-pregnant | <i>Menstrual cycles of attempt time at study entry</i> |            |
|-----------|------|------------|--------------|--------------------------------------------------------|------------|
|           |      | n (%)      | n (%)        | 0                                                      | 1          |
| Model I   | 3195 | 2747 (86%) | 448 (14%)    | 1348 (42%)                                             | 1847 (58%) |
| Model II  | 3476 | 2406 (69%) | 1070 (31%)   | 1462 (42%)                                             | 2014 (58%) |
| Model III | 4133 | 2747 (66%) | 1386 (34%)   | 1737 (42%)                                             | 2396 (58%) |
| Model IV  | 1571 | 1320 (84%) | 251 (16%)    | 663 (42%)                                              | 908 (58%)  |
| Model V   | 1722 | 1139 (66%) | 583 (34%)    | 726 (42%)                                              | 996 (58%)  |
| Model VI  | 1957 | 1333 (68%) | 624 (32%)    | 819 (42%)                                              | 1138 (58%) |

Note: Model I predicts pregnancy in <12 menstrual cycles; Model II predicts pregnancy in <7 menstrual cycles; Model III predicts the probability of pregnancy within each menstrual cycle for up to 12 cycles of follow-up; Model IV predicts pregnancy in <12 menstrual cycles among nulligravid women with no history of infertility; Model V predicts pregnancy in <7 menstrual cycles among nulligravid women with no history of infertility; Model VI predicts the probability of pregnancy within each menstrual cycle for up to 12 cycles of follow-up among nulligravid women with no history of infertility.

**Table SVII. Distribution of each class by the number of menstrual cycles of attempt time at study entry.**

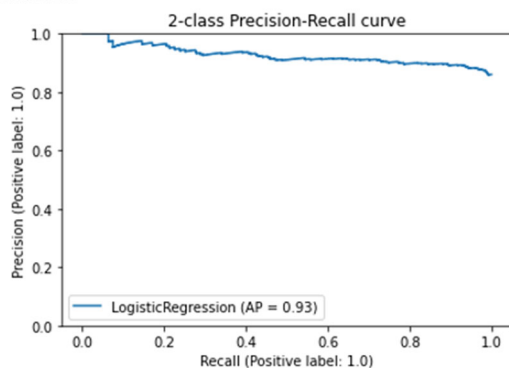
|           | <i>No menstrual cycles of attempt time at study entry</i> |              | <i>One menstrual cycle of attempt time at study entry</i> |              |
|-----------|-----------------------------------------------------------|--------------|-----------------------------------------------------------|--------------|
|           | Pregnant                                                  | Non-pregnant | Pregnant                                                  | Non-pregnant |
| Model I   | 1205 (38%)                                                | 143 (4%)     | 1542 (48%)                                                | 305 (10%)    |
| Model II  | 1086 (31%)                                                | 376 (11%)    | 1320 (38%)                                                | 694 (20%)    |
| Model III | 1213 (29%)                                                | 524 (13%)    | 1557 (38%)                                                | 839 (20%)    |
| Model IV  | 588 (37%)                                                 | 75 (5%)      | 732 (46%)                                                 | 176 (11%)    |
| Model V   | 518 (30%)                                                 | 208 (12%)    | 621 (36%)                                                 | 375 (22%)    |
| Model VI  | 591 (30%)                                                 | 228 (12%)    | 742 (38%)                                                 | 396 (20%)    |

Note: Model I predicts pregnancy in <12 menstrual cycles; Model II predicts pregnancy in <7 menstrual cycles; Model III predicts the probability of pregnancy within each menstrual cycle for up to 12 cycles of follow-up; Model IV predicts pregnancy in <12 menstrual cycles among nulligravid women with no history of infertility; Model V predicts pregnancy in <7 menstrual cycles among nulligravid women with no history of infertility; Model VI predicts the probability of pregnancy within each menstrual cycle for up to 12 cycles of follow-up among nulligravid women with no history of infertility.

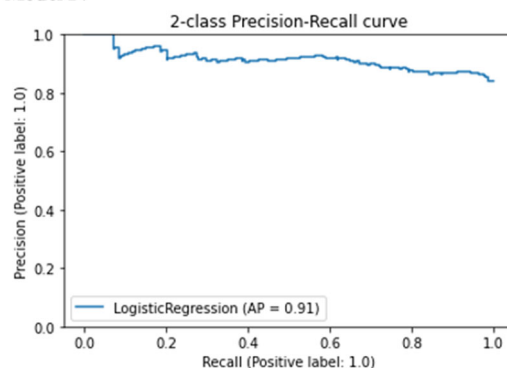
**Figure SI. Area under the precision-recall curves (AUPRC) for Models I, II, IV, and V.**

Note: Model I predicts pregnancy in <12 menstrual cycles; Model II predicts pregnancy in <7 menstrual cycles; Model IV predicts pregnancy in <12 menstrual cycles among nulligravid women with no history of infertility; Model V predicts pregnancy in <7 menstrual cycles among nulligravid women with no history of infertility.

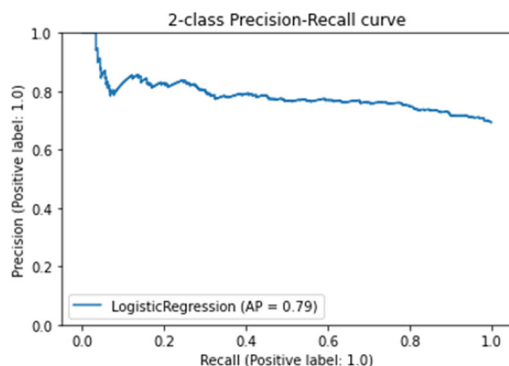
**Model I**



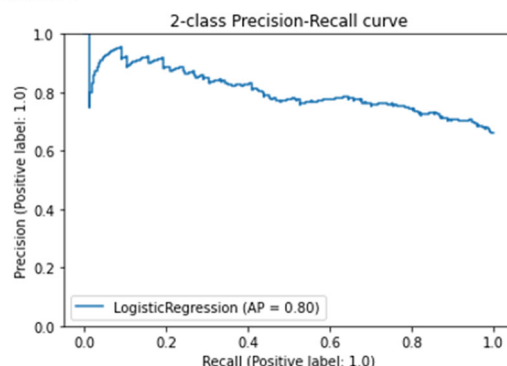
**Model IV**



**Model II**

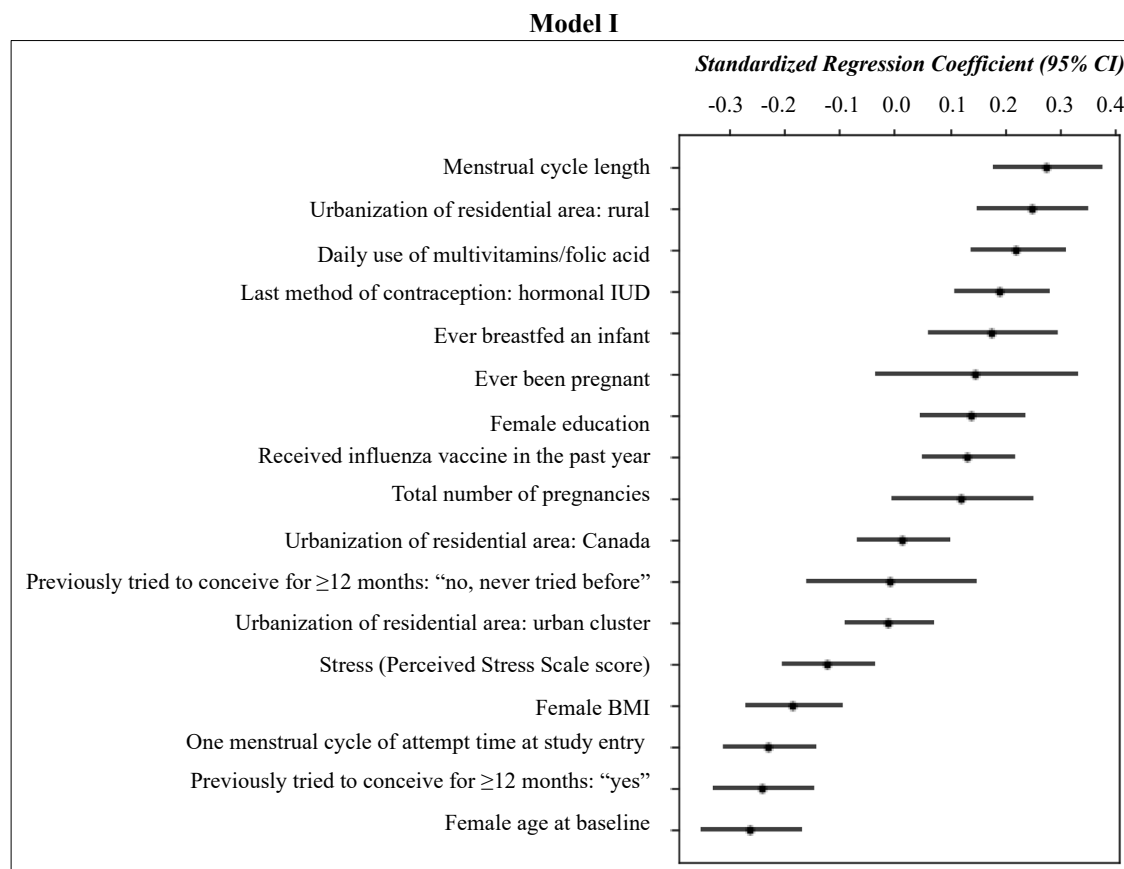


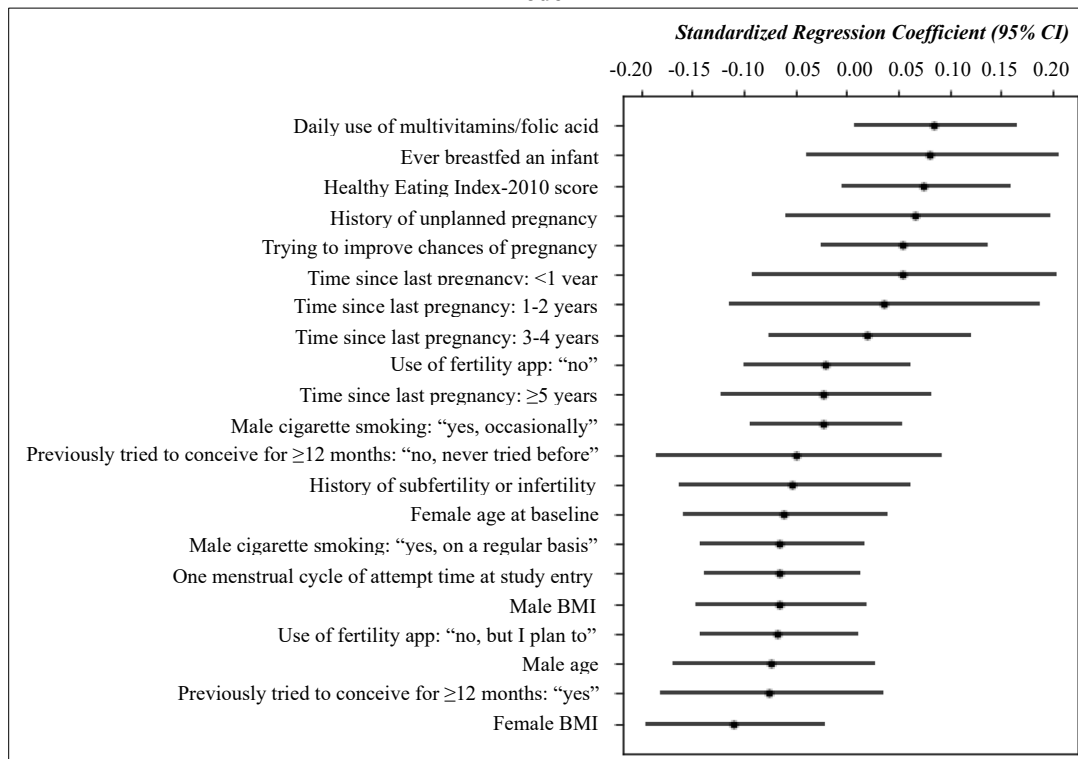
**Model V**



**Figure SII. Visualization of the model coefficients and their 95% confidence intervals.**

Note: These plots present the model coefficients with error bands equivalent to 95% confidence intervals. The variables are ordered according to their mean coefficient. All plots are associated with the parsimonious L2LR version of models, consistent with the rest of the Manuscript. Model I predicts pregnancy in <12 menstrual cycles; Model II predicts pregnancy in <7 menstrual cycles; Model III predicts the probability of pregnancy within each menstrual cycle for up to 12 cycles of follow-up.



**Model II****Model III**