I. Overview

As part of our mission to create noninvasive, accurate diagnostic tools for individuals with endometriosis, our team developed a predictive model for binary classification of whether or not a woman has endometriosis based on a set of clinical and demographic variables. Our model can be used as a diagnostic tool as well as to discover the relative importance of potential endometriosis risk factors. When compared to other models currently available for endometriosis diagnosis, our model was the most accurate, least expensive, and most inclusive. To increase accessibility of our model, we integrated our model into a software tool that clinicians can use to assess patients' risks of endometriosis.

II. Background Research

Endometriosis is a disease characterized by growth of endometrial-like tissue outside the uterine cavity (Giudice, 2010). This chronic disease affects 10% of women worldwide with symptoms including incapacitating pelvic pain, abnormal uterine bleeding, and infertility (Eskenazi & Warner, 1997). Currently, the only method of diagnosis is laparoscopy, a surgical procedure used to view the abdomen and pelvic area (Kennedy et al., 2005). However, the extent of disease identified through laparoscopy is often poorly correlated with severity of symptoms (Argarwal et al., 2019). This procedure also has technical limitations as it relies on visualization of endometriotic lesions which can be complicated by lesion appearance and location (Agarwal et al., 2019). Additionally, the surgical procedure is a more costly type of diagnostic and therefore increases healthcare-related costs for women with endometriosis (Soliman et al., 2017). Even with this invasive diagnostic tool available, it takes approximately 11 years between the onset of symptoms and diagnosis of endometriosis (Agrawal et al., 2019, Hudelist, et al., 2012, Staal et al., 2016). This diagnostic delay and lack of non-invasive techniques is attributed to lack of research, awareness, and medical education on the pathology and symptoms of endometriosis (Agrawal et al., 2019).

Symptom-based predictive models have been created as a diagnostic tool for a wide variety of diseases and conditions such as breast cancer (Armer et al., 2003), preterm labor (Carter et al., 2020), and even COVID-19 (Ronald et al., 2020). Therefore, our team decided to investigate whether a predictive model that utilizes clinical and demographic data could be created as a diagnostic tool for endometriosis. In creating this model, we aim to not only address the lack of non-invasive diagnostics for endometriosis but also increase physician awareness of the disease and risk factors that are associated with it by creating an accessible software tool that guides physicians through the diagnostic process.

Risk factors for endometriosis have been well-studied through epidemiological studies. Our model was trained with the factors included in endometriosis researcher Dr. Idhaliz Flores' patient registry, which includes the most relevant risk factors previously identified to have strong correlations with endometriosis. These factors include:

- Symptoms such as incapacitating pain, dyspareunia (pain during sexual intercourse), and dysmenorrhea (pain during menstruation)
- Lifestyle factors such as height, weight, body mass index, and exercise habits
- Reproductive health factors such as age of menarche, period length, menstrual cycle length, and infertility

- Comorbidities such as autoimmune diseases, pelvic/gynecological conditions, ovarian cysts, and musculoskeletal pain
- Family history of endometriosis

While predictive models for endometriosis have been constructed before using similar data on symptoms, demographics, family history, and medical history, none have generated enough predictive power to be accepted as a valid diagnostic tool (Chapron et al., 2005, Eskenazi et al., 2001, Nnoaham et al., 2012). In fact, a recent article that identified 16 endometriosis predictive models concluded that none were sufficient for clinical use (Surrey et al., 2017). A contributing factor to the lack of success in an effective clinical model is the large diagnostic bias seen throughout endometriosis research through inconsistent diagnoses. The definition of endometriosis is inconsistent worldwide, and there is not a consensus regarding endometriosis symptoms. Additionally, there is a disparity of endometriosis diagnoses by racial and ethnic groups despite no evidence that the true rates of endometriosis are associated with race (Nnoaham et al., 2009, Missmer et al. 2004). These diagnostic biases affect all epidemiologic studies on the rates of endometriosis and the associated risk factors. Additionally, while many models have been made to analyze associations between specific symptoms and endometriosis, most previous models do not consider multiple predictor variables, where a combination of variables are used to determine whether or not a woman has endometriosis (Apostolopoulos et al., 2016, Droz & Howard, 2011, Fuldeore & Soliman, 2017, Peterson et al., 2013, Schliep et al., 2015). This makes it necessary to create a framework for a predictive model that can screen for risk and contribute to more accurate diagnostic tools.

Our model improves upon previous work and can be used as a clinically relevant predictive model for endometriosis. Previous models were constructed with data from symptomatic cases in order to predict the need for surgery to diagnose and subsequently treat endometriosis. However, the patients in our proposed study were confirmed to have or not have endometriosis through surgery and subsequent histopathological confirmation, which will improve our model accuracy in the prediction of whether or not women have endometriosis. Additionally, we seek to improve the accessibility of our predictive clinical model by integrating it into website format, reducing the need for any interviewer-administered questionnaires.

III. Ethics and Institutional Review

Prior to receiving the patient data, we submitted our data collection, storage, and analysis protocol to the University of Rochester's Institutional Review Board (IRB) for approval. Our IRB determined that since we were obtaining de-identified datasets with all dates removed, our work did not count as human subjects research. Therefore, our IRB provided a waiver for this study. Additionally, all students, teaching assistants, and advisors completed CITI training for Biomedical, Social, Educational, and Behavioral Research.

IV. Study Sample

Data collected by Dr. Idhaliz Flores' Endometriosis Research Lab at Ponce Health Sciences University - Ponce Research Institute from 2001 to 2010 was de-identified and then transferred to our team. This set contained data from 1,560 women from Puerto Rico, including 1,182 with surgically and

histopathologically confirmed endometriosis and 378 surgically confirmed negative controls. None of the women in the dataset had malignant conditions.

V. Data Cleaning & Variable Selection

Our team aims to develop diagnostic tools that are simple to use. This includes decreasing the number of inputs in our model for ease of use by physicians and patients. To accomplish this, we cleaned the data as follows before building our model, effectively reducing the number of variables from 202 to a final dataset of 48 predictors (*Figure 1*).

Predictors Included in Model 1	
Demographics	Age, Educational background, Health insurance type
Reproductive Health	Age of menarche, Menstrual cycle length, Period length, Regular cycle (Y/N), Number of children, Number of miscarriages, Problems getting pregnant (Y/N), Age when 1st child was born, Prior oral contraceptive use (Y/N) Current oral contraceptive use (Y/N), Infertility (Y/N), Condom use (Y/N), Rhythm contraceptive use (Y/N), Infertility treatment (Y/N), Age that symptoms began
Symptoms	Dysmenorrhea (Y/N), Incapacitating pain (Y/N), Dyspareunia (Y/N), Musculoskeletal symptoms (Y/N), Abnormal uterine bleeding (Y/N), Pelvic pain (Y/N), Leg pain (Y/N), Dizziness (Y/N), Asymptomatic (Y/N)
Life Style Factors	Smoking status, NSAID use (Y/N), BMI, Times exercise per week
Prior Procedures and Examinations	Abnormal Pap (Y/N), Tube ligation (Y/N)
Family History (Y/N)	
Conditions	Gastrointestinal conditions (Y/N), Allergies, (Y/N), Breast cysts (Y/N), Uterine Fibroids (Y/N), Ovarian cysts (Y/N), Gynecological infections (Y/N), Gynecological or pelvic conditions (Y/N),Diabetes (Y/N), Cardiovascular conditions (Y/N), Autoimmune disease (Y/N), Insomnia (Y/N), Chronic infections (Y/N)

Figure 1: Predictors included in the first model iteration (Model 1)
NSAID = Nonsteroidal anti-inflammatory drugs; Pap = Papanicolaou test, BMI = body mass index

1. **Variables with >90% of "unknown" responses were removed.**
   We removed variables with large amounts (>90%) of missing data. We checked whether these variables contributed to the model by analyzing each variables' mean decrease accuracy (MDA) which measures the average decrease in model accuracy when the variable is removed.

Variables with > 90% “unknown” had MDAs equal to 0.000 indicating that they were neither increasing nor decreasing model accuracy and therefore were not contributing to the model. We removed 51 variables based on this criteria.

2. **Removed highly correlated variables.**
   In order to reduce the number of predictors in our model without losing model accuracy, and to enhance our model performance, we tested each possible pairing of variables to determine which variables were strongly correlated. If a pair of variables had a correlation coefficient >0.7 or < -0.7 (determined by cor() function in R programming language), then one of the variables in the pair was removed. We determined which variable to keep based on which variable was easier to collect or understand, ensuring that our model was as simple and convenient for both patients and physicians. We removed five variables based on this criteria.
3. **If a variable had responses that could only be true for patients with endometriosis, then the variable was excluded.**
   If we included variables that could only be true for patients with endometriosis, such as endometriosis severity, then this would bias the model. It would be impossible for someone to know the severity of their endometriosis prior to receiving a diagnosis, and therefore these variables, and those similar to it, should not be included in diagnostic models for endometriosis. We removed 12 variables based on this criteria.
4. **Symptoms and conditions were condensed into broad categories.**
   A wide variety of symptoms and conditions were reported in this study. In fact, there were 115 unique symptoms and conditions across all subjects. Therefore, instead of counting each sign and condition as a unique predictor, we condensed them into broad categories such as “autoimmune diseases,” “gastrointestinal conditions,” etc. We reduced the number of predictors by 93 variables based on this criteria.
5. **Specific family relationships between subjects and their family members with endometriosis were condensed into a single variable that captures family history.**
   Similar to criteria (4), we reduced the various relationships between subjects and family members with endometriosis into a single variable to increase the ease of use of the model. We reduced the number of predictors by five variables based on this criteria.

The survey that was used to collect the patients’ data, and therefore all variables in the raw dataset, can be found here.
Survey

VI. Modeling Framework

In order to create our predictive model, we decided to use a machine-learning algorithm known as random forest. Random forest algorithms work by creating **multiple decision trees** to classify binary outputs. When using a decision tree for binary classification, the data is split based on various parameters until it comes to a final decision on how to classify the data (*Figure 2*). Many decision trees can be made from a single dataset. For example, the predictor(s)’ threshold value, which is used to split the data, can be different for each tree. Additionally, since variables are randomly sampled at each split, the order the predictor variables appear in the decision tree is different in each tree. The random forest algorithm takes advantage of the variety of decision trees that can be constructed by a single dataset. Rather than using

one single decision tree to classify data, random forest uses the majority classification determined by multiple decision trees (hence, a forest). Usually, the more trees in your forest, the more accurate your random forest algorithm is.

In our model, 500 decision trees are used, and 17 predictors are randomly sampled at each split. Each of the 500 decision trees classify the patients as having endometriosis or not having endometriosis based on a unique order of predictors and predictor threshold values. However, the final classification is based on the majority classification across the 500 trees. For example, if five decision trees classify a patient with endometriosis but the other 495 decision trees in the forest classify the patient as healthy, then our model would predict the patient is healthy.

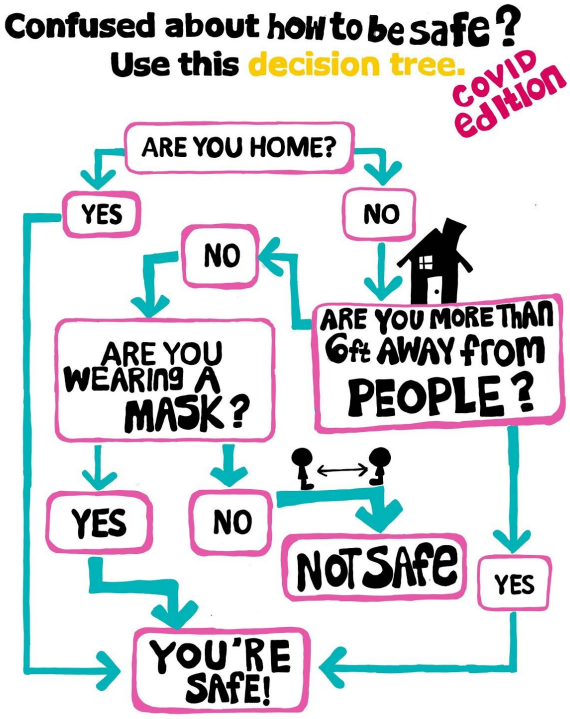


Figure 2: Example Decision Tree
This is a single decision tree that classifies a person as "safe" or "not safe" from COVID-19 based on the following predictors: whether or not they are home, wearing a mask, and/or standing 6 ft apart from people. In a random forest, the order of these decisions can change.

Model Parameters

The predictive model was developed using the tidyverse and randomForest libraries in the R programming language.

We built our model using 500 decision trees (ntree), 17 predictor variables randomly sampled at each split (mtry), and importance set to true (T) and importance type set to one. Setting importance equal to true and type equal to one enables us to analyze the mean decrease accuracy for each predictor in our model.

Training, Validation, and Model Performance Evaluation

We built our model on a training set consisting of a randomly selected 70% subset of our study sample and then validated it on the remaining 30%.

To evaluate our model, we utilized confusion matrices, accuracy, sensitivity, and specificity (*Figure 3 & 4*).

Statistic	Description
False Positive	Number of negative controls incorrectly classified as having endometriosis
False Negatives	Number of subjects with endometriosis incorrectly classified as not having endometriosis
True Positive	Number of negative controls correctly classified as not having endometriosis
True Negatives	Number of subjects with endometriosis correctly classified as not having endometriosis
False Negative Rate	The proportion of negative controls incorrectly classified as having endometriosis
False Positive Rate	The proportion of subjects with endometriosis incorrectly classified as not having endometriosis
Sensitivity	The proportion of endometriosis subjects that our model correctly identified as having endometriosis
Specificity	The proportion of negative controls that our model correctly identified as healthy
Accuracy	The proportion of study subjects correctly classified as either having or not having endometriosis

Figure 3: Summary of statistics used for model performance evaluation

	True Negative	True Positive
Predicted Negative	Number of True Negatives	Number of False Negatives
Predicted Positive	Number of False Positives	Number of True Positives

Figure 4: Overview of a Confusion Matrix

VII. Initial Model Results

Our cleaned dataset (n = 1560) was randomly split into a training set (n =1092) and validation set (n = 468). This model (Model 1) was then tested on the validation dataset and found to have a sensitivity of 88.76%, specificity of 68.75% and accuracy of 83.97% (*Figure 5*).

A.

	True Negative	True Positive
Predicted Negative	77	40
Predicted Positive	35	316

B.

Model 1 Results	
Accuracy	83.97%
Sensitivity	88.76%
Specificity	68.75%

Figure 5: Model 1 Results
(A) Confusion matrix for Model 1 on validation set (n = 468).
(B) Accuracy, sensitivity, and specificity of Model 1 on validation set (n = 468).

VIII. Improving the Model

After observing that our model had a higher false positive rate (~31%) than false negative rate (~11%), we hypothesized that this could be due to building the model on an unbalanced patient dataset containing more endometriosis patients (n=1182) than healthy controls (n=378). We decided to create a new dataset that included all 378 controls and a randomly selected pool of 378 endometriosis patients. We then re-built the model (Model 2) on a training set made from a random sample of the new, balanced dataset.

This decreased our error rate from **14%** to **7.4%** when tested on the validation dataset (*Figure 6*).

A.

	True Negative	True Positive
Predicted Negative	107	11
Predicted Positive	6	103

B.

Model 2 Results	
Accuracy	92.51%
Sensitivity	90.35%
Specificity	94.69%

Figure 6: Model 2 Results
(A) Confusion matrix for Model 2 on validation set (n = 227).
(B) Accuracy, sensitivity, and specificity of Model 2 on validation set (n = 227).

One of our main goals was to determine which variables were most important in predicting endometriosis in our model and to then compare the most important predictors to what is already known about symptoms associated with endometriosis. Using the importance() function with type = 1, we determined how each variable was affecting our model through a metric known as Mean Decrease Accuracy (MDA). MDA measures the average decrease in model accuracy when the variable is removed. The more positive the MDA value, the more important the variable is in the model.

After constructing an importance plot using Model 2 (*Figure 7*), the two most important predictors stood out. First, the age that endometriosis symptoms began not only had the largest MDA value (65.57) but was over two-times the value of the second most important predictor (28.21). Similarly, the second most important predictor, health insurance type, had an MDA value of about 1.5 times the third-highest predictor (17.1). In our dataset, health insurance type was categorized as “Private,” “Public,” or “None.” This is of interest to our team as, to our knowledge, an association between endometriosis and age symptoms present and endometriosis and health insurance type have not been previously described.

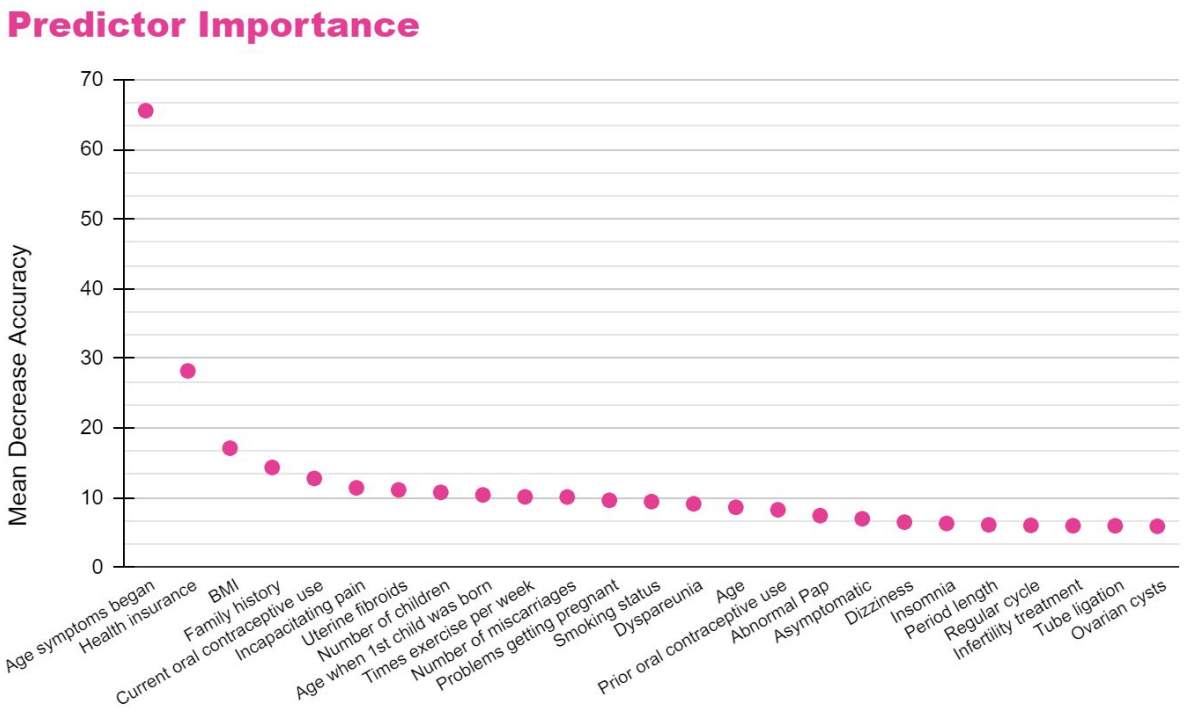


Figure 7: Predictor Importance Model 2

Given this surprising result, we further investigated the potential relationships between these predictors and endometriosis. However, we realized that these results were likely artifacts as our controls had 87% unknown for age of symptoms while our cases had 90% unknown for health insurance. Therefore, we decided to remove these two variables as well as any other variable where either the cases or control had > 85% unknown responses. This left us with 42 variables (*Figure 8*).

A.

Predictors with > 85% in Unknown or Control		
BMI	Infertility Treatment (Y/N)	Health insurance type
Current oral contraceptive use (Y/N)	Times exercise per week	Age that symptoms began

B.

Predictors Included in Final Model	
Demographics	Age, Educational background
Reproductive Health	Age of menarche, Menstrual cycle length, Period length, Regular cycle (Y/N), Number of children, Number of miscarriages, Problems getting pregnant (Y/N), Age when 1st child was born, Prior oral contraceptive use (Y/N), Infertility (Y/N), Condom use (Y/N), Rhythm contraceptive use (Y/N)
Symptoms	Dysmenorrhea (Y/N), Incapacitating pain (Y/N), Dyspareunia (Y/N), Musculoskeletal symptoms (Y/N), Abnormal uterine bleeding (Y/N), Pelvic pain (Y/N), Leg pain (Y/N), Dizziness (Y/N), Asymptomatic (Y/N)
Life Style Factors	Smoking status, NSAID use (Y/N)
Prior Procedures and Examinations	Abnormal Pap (Y/N), Tube ligation (Y/N)
Family History (Y/N)	
Conditions	Gastrointestinal conditions (Y/N), Allergies, (Y/N), Breast cysts (Y/N), Uterine Fibroids (Y/N), Ovarian cysts (Y/N), Gynecological infections (Y/N), Gynecological or pelvic conditions (Y/N),Diabetes (Y/N), Cardiovascular conditions (Y/N), Autoimmune disease (Y/N), Insomnia (Y/N), Chronic infections (Y/N)

Figure 8: Predictors included in third model iteration (Model 3)
(A) Predictors with > 85% unknown in either cases or control.
(B) Predictors included in the third model iteration. NSAID = Nonsteroidal anti-inflammatory drugs; Pap = Papanicolaou test

Using our 42-variable dataset, we re-built the model (Model 3) on a balanced dataset following the same principles previously described. This decreased the accuracy from 93% to 85% (*Figure 9*). However, despite the decrease in accuracy, Model 3 is the best model as it was built on a more complete dataset than the dataset used in Model 2. Predictors with significantly more unknowns in one group versus another (*Figure 8A*), enable the model to use "unknowns" to classify the women. While using unknown is appropriate in cases where there are clinical or scientific explanations for individuals in a specific group to have more unknowns than in another group, this was not the case with the six variables we removed from our dataset. Therefore, even though a decrease in accuracy was observed, Model 3 is more likely to perform with consistent accuracy than Model 2. Additionally, although Model 3 has similar accuracy to Model 1, Model 3 has a 17% improvement in specificity when compared to Model 1.

A.

	True Negative	True Positive
Predicted Negative	97	18
Predicted Positive	16	96

B.

Final Model Results	
Accuracy	85.02%
Sensitivity	84.21%
Specificity	85.84%

Figure 9: Final Model (Model 3) Results
(A) Confusion matrix for Final Model on validation set (n = 227).
(B) Accuracy, sensitivity, and specificity of Final Model on validation set (n = 227).

IX. Relative Importance of Clinical Symptoms

The top 25 most important predictors (*Figure 10*) align with previously reported associations with endometriosis (Agarwal et al., 2019, Cramer & Missmer, 2002., Matalliotakis et al., 2008, Mcleod & Retzloff, 2010, Missmer et al., 2004, Tanbo & Fedorcsak, 2017). Our top three most important variables are experiencing incapacitating pain, having a family history of endometriosis, and problems getting pregnant, respectively. This result agrees with the current understanding of endometriosis risk factors as severe pelvic pain, genetic relationships, and infertility have been widely studied in relation to endometriosis, and strong associations between these variables and the disease have been described (Agarwal et al., 2019, Matalliotakis et al., 2008, Tanbo & Fedorcsak, 2017).

While not the most important predictor, smoking status did appear to contribute to our model accuracy. This adds to the current debate on whether or not there is an association between smoking and endometriosis risk. Studies have demonstrated a relationship between cigarette smoke and elevated estrogen levels in the lungs (Cohen & Smith, 2016), leading researchers to wonder if cigarette smoke could contribute to the elevated levels of estrogen seen in women with endometriosis. While some studies have found an association between endometriosis and smoking (Garavaglia et al., 2017), others have not been able to validate such results (Hemert et al., 2019). The findings that smoking status did contribute to our model emphasizes the need for further research to decipher whether or not there is a physiological relationship between cigarette smoke and endometriosis.

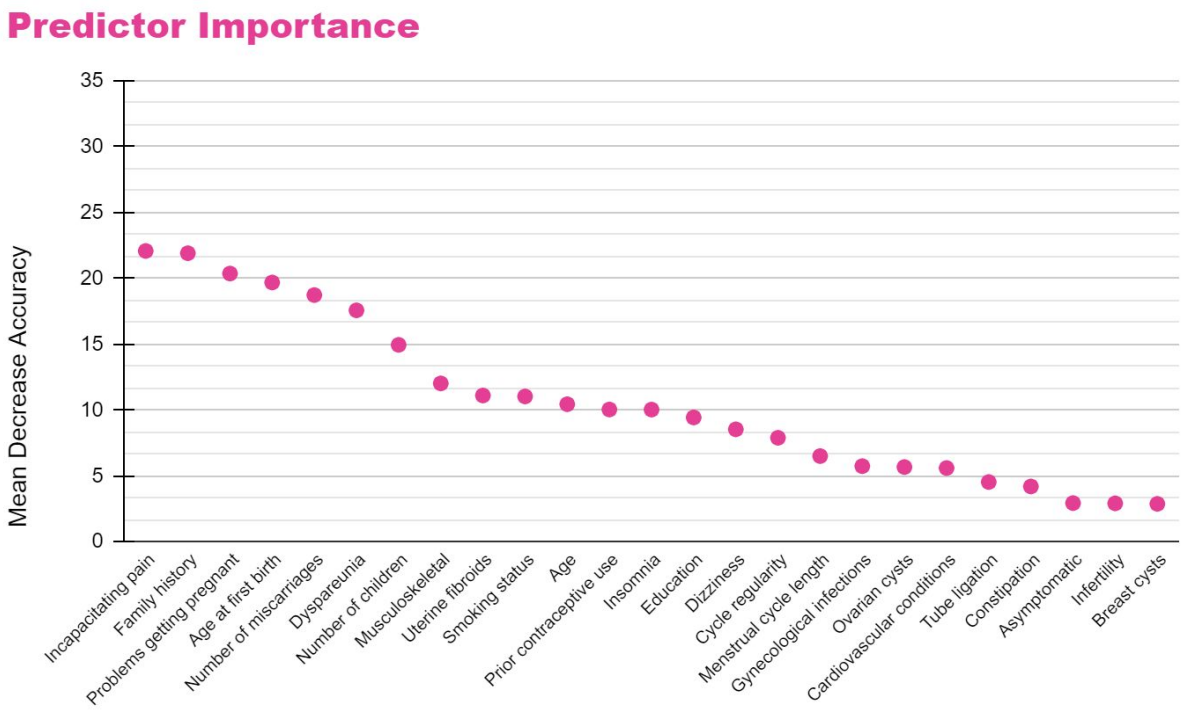


Figure 10: Predictor Importance for Top 25 Predictors, Final Model

X. Investigating Model Accuracy by Age of Presentation

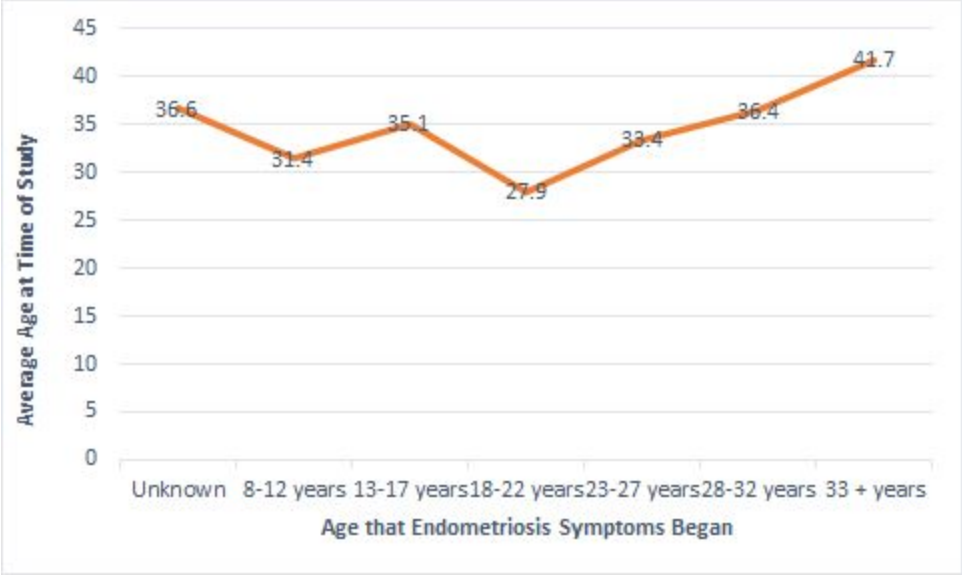
We were interested in asking whether endometriosis prevalence and our model performance varies depending on the age of onset of symptoms (*Figure 11*).

It appears that the accuracy of our model decreases as age of presentation increases (*Figure 11B*). We thought that this could be due to women with older ages of presentation being older than the women with younger ages of presentation at the time of the study. Older age could contribute to differences in sex hormone levels that would alter reproductive health factors or lead to increase in comorbidity diagnosis. Therefore, we hypothesized that these factors associated with aging were contributing to the decrease in model accuracy. However, the average age of women did not increase as the age of presentation increased, and therefore this is likely not the case (*Figure 11A*).

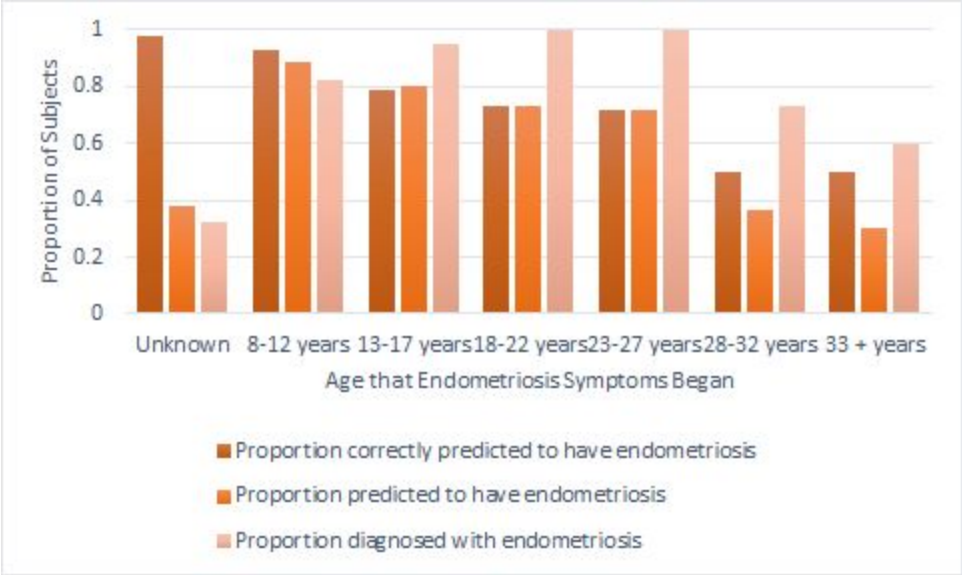
Another possibility is that women with a younger age of presentation have, on average, distinct responses to our top three predictors when compared to women with an older age of presentation. Therefore, we analyzed each age of symptom presentation groups based on the proportion of women in each age range with positive responses to incapacitating pain, family history of endometriosis, and problems getting pregnant (*Figure 11C*). Interestingly, there is a higher proportion of women with positive responses to the top three predictors between the age of presentation at 8-22 years old than the age of presentation at 23 years old or older. In fact, the women in the older age of presentation groups had positive response proportions that were more similar to the unknown age group, which contained 68% negative controls.

All together, it appears that the women in the older age of presentation groups have responses to that top three predictors that are more similar to the controls than the women in the younger age of presentation groups. These differences in proportion of positive responses could account for the differences in model accuracy between the groups.

A.



B.



C.

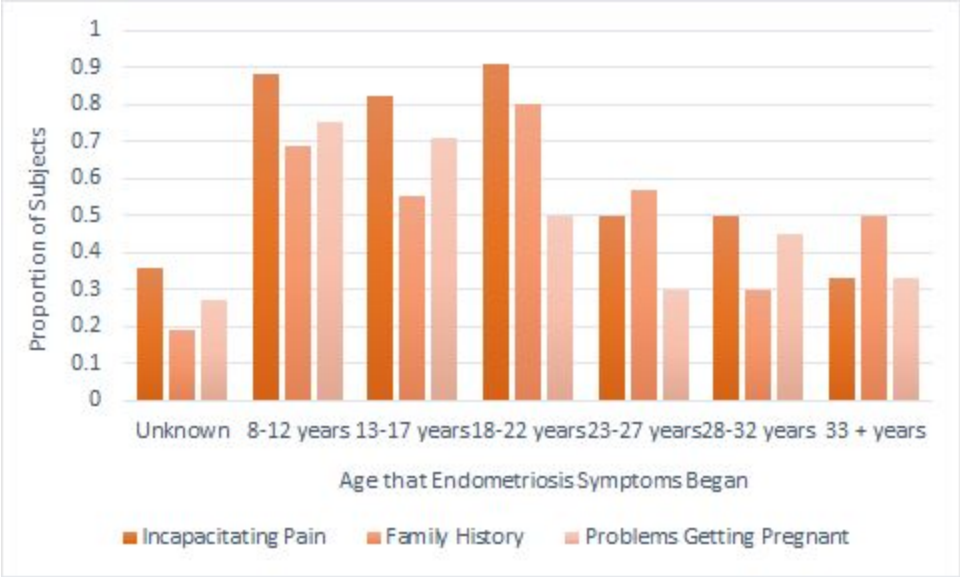


Figure 11: Analyzing model results by age group

XI. Comparing our Model to Current Endometriosis Clinical Models

In order to see how our model compares to current endometriosis predictive models, our team reviewed endometriosis predictive models described in the literature. To be included in our review, the study had to meet the following criteria:

1. **Not study a specific site of endometriosis.**
   Some predictive models focused on predicting site-specific endometriosis, such as the bladder (Fedele et al., 2007), rectovaginal (Griffiths et al., 2007), or deep infiltrating endometriosis (Lafay Pillet et al., 2014, Perello et al., 2017). We decided not to compare our model to these models as our dataset included women with all forms of endometriosis, and it was not our goal to diagnose site-specific endometriosis.
2. **Include both fertile and infertile women in the study.**
   While infertility does affect approximately 30-50% of women with endometriosis (Bulletti et al., 2010), we did not limit our study sample to exclusively women with infertility. However, some studies do limit their sample to infertile women (Ashrafi et al., 2016, Heitman et al., 2014, Meuleman et al., 2009, Paulson, 2009). We excluded these studies since our dataset utilizes infertility as a predictor rather than an inclusion criteria.
3. **Report results of the overall model including sensitivity, and specificity.**
   Many of the models we reviewed were most interested in determining associations between variables and endometriosis, rather than investigating the use of variables as predictors in an overall diagnostic tool (Apostolopoulos et al., 2016, Droz & Howard, 2011, Fuldeore & Soliman, 2017, Peterson et al., 2013, Schliep et al., 2015). Without statistics such as sensitivity and specificity, it is difficult to make performance comparisons between models, and it is unclear whether the authors intended for their work to be a diagnostic tool or to inform associations between symptoms and endometriosis.

4. **Indicate variables used in the final model.**
   In order to effectively compare models, it was necessary to know what clinical or non-clinical variables were used to build each model. This enables us to compare the importance of the same predictors across models, investigate how the predictors used may affect model accuracy, and provides insight into the ease of use and overall cost of current predictive models. As our main goal was to create a non-invasive, inexpensive, and easy to use diagnostic, we did not utilize predictors such as ultrasound imaging in our model, which would increase the healthcare utilization, cost, and complexity of this diagnostic tool. It was important for us to know whether other models required this information in order to achieve adequate sensitivity and specificity.

After reviewing 14 studies, we identified one model that met the above criteria to critically compare to our model (Nnoaham et al., 2012).

There are key differences between the our model and the model reported in Nnoaham et al., 2012 (*Figure 12)*. The first, and possibly most important, distinction is that our predictive model was developed and validated on datasets that included women with all stages (minimal, mild, moderate, and severe) of endometriosis. However, in Nnoaham et al., the model was developed (n=771) and validated (n=625) on patients with moderate to severe endometriosis only. Nnoaham et al. did attempt to create a model across all stages of endometriosis but the specificity of the model was only 43.8% in the validation cohort. This could be due to slight differences in inclusion and exclusion criteria between the two studies. While our study included all women with histopathologically confirmed endometriosis patients and surgically confirmed controls, the Nnoaham study utilized visual laparoscopy diagnostics to determine whether or not a study subject had endometriosis. Similarly, women on oral contraceptives within the past 3 months or with amenorrhea in the last 3 months were excluded from the Nnoaham study but not from our study. Since oral contraceptives are commonly used to prevent pregnancy and are a treatment for endometriosis symptoms, including women taking oral contraceptives in our study increases the applicability of our model into the real world. Additionally, our dataset contained 26 more variables than Nnoaham (*Figure 13*). Nnoaham, et al. also utilized ultrasound imaging to increase the accuracy of their model, while our model does not require this or any other imaging tool therefore making our diagnostic less expensive and more convenient for patients.

	Nnoaham et al., 2012	**Team UteRus**
Classification Method	Logistic Regression	Random Forest
Sample Size **A. Training Set** **B. Validation Set**	1396 A. 771 B. 625	756 A. 529 B. 227
Accuracy	84.90%	85.02%
Sensitivity	82.26%	84.21%

Specificity	89.93%	85.84%
Inclusion Criteria	Women with moderate to severe endometriosis and healthy negative controls determined by visual laparoscopy diagnostics	Women with histopathologically confirmed endometriosis patients and surgically confirmed negative controls
Exclusion Criteria	Women on oral contraceptives or with amenorrhea in the last 3 months	Women with malignant conditions
Number of Predictors	17	42
Only moderate to severe endometriosis detected?	Yes	No
Imaging used?	Yes, ultrasound	No

Figure 12: Comparing Models Between Nnoaham et al., 2012 and Team UteRus

Variables included in Nnoaham et al., 2012
Average cycle length
Pelvic pain during periods
Pain on opening bowels during period
Pain usually approximately same time in cycle (just after periods)
IBS
IBS-M
Frequency of nocturia in general*
Number of live births
Ever diagnosed blocked tubes as a reason for subfertility*
Ovarian cysts (benign) ever diagnosed
Polycystic Ovarian Syndrome ever diagnosed

Asthma ever diagnosed
Any atopic condition ever diagnosed*
Family history of prostate cancer in first- or second-degree relative*
Asian or Oriental ethnicity*
Black ethnicity*

Figure 13: Variables included in Nnoaham et al., 2012 predictive model
*predictor not included in Team UteRus model

Overall, there appears to be a lack of research focused on creating non-invasive clinical predictive models for endometriosis. Based on our criteria, we were only able to identify one other model which was roughly as accurate as our model. However, it utilized ultrasound imaging to increase model specificity. **Our model does not utilize ultrasound imaging, thus making our model less expensive and more convenient for patients.** Additionally, the previous model was only able to sensitively and specifically classify women with moderate to severe endometriosis while **our model can more sensitively and specifically classify women with all stages of endometriosis, including mild, minimal, moderate, and severe.** We, therefore, conclude that our model is the most accurate, least expensive, and most inclusive model for endometriosis that is currently available.

XII. Future Directions

After creating this model and reviewing the literature, our team has thought about the potential next steps we could take to improve our model and further understand how our model fits into researchers' current understanding of endometriosis predictors.

Investigate age of symptom presentation and health insurance association with endometriosis

It is important to note that because random forests create multiple decision trees (as described in Section V: Modeling Framework), it is difficult to determine linear relationships and correlations between predictor variables and the class type (endometriosis vs. healthy) because each predictor is likely used in a unique way in each decision tree. Our team would conduct further statistical analyses to elucidate the association between endometriosis and age of presenting symptoms as well as endometriosis and health insurance type.

Alter the survey and obtain a new validation cohort

While creating our model, we noticed that some predictors had more unknown responses than others. We believe that this could be due to the way the questions were presented in the survey and that altering the question format could lead to more responses. As previously discussed, we had to exclude variables with clinical significance from our model due to a high percentage of unknowns within one or both of our groups (*Figure 8*). For example, the age that symptoms began had an unknown rate of 87% within the control group. The question was formatted such that subjects write in the age. Perhaps many women do

not know the exact age the symptoms began and, as such, leave this question blank. If another validation cohort were obtained, we could alter the question to instead be multiple choice with age-ranges provided so women can be more confident in responding. All questions in the survey should be evaluated for similar changes in format.

Validate the model on a more geographically, ethnically, and racially diverse cohort

We built our model using data from Puerto Rico, where a vast majority of our subjects were of Latinx descent. Therefore, it is possible that the high accuracy achieved is in part due to lack of geographic, ethnic, and racial diversity in our sample and that our model would not perform as well on a more diverse sample. Our team's next steps would be to validate our model on a more diverse sample that better represents all ethnic and racial backgrounds and, if necessary, re-train the model.

XIII. Summary

Our team used a random forest algorithm to create a predictive model for endometriosis based on 49 clinical and demographic variables. We demonstrated that our final model is 85% accurate with a sensitivity of 84% and specificity of 86% and that incapacitating pain, family history of endometriosis, and problems getting pregnant are the three most important predictors in our model. When compared to current endometriosis predictive models, our model is the **most accurate, least expensive, and most inclusive model.**

Additionally, we integrated our predictive model into a software tool that physicians can utilize to not only non-invasively and accurately diagnose endometriosis but to also become more familiar with clinical symptoms and demographics that can predict the disease.

A limitation of our model is the lack of geographic, ethnic, and racial diversity as the study sample solely contained women from Puerto Rico. It is possible that our model accuracy decreases when validated on a more diverse cohort. Therefore, future work should focus on re-training and validating our model on a sample that is geographically, ethnically, and racially diverse.

References

Agarwal, S. K., Chapron, C., Giudice, L. C., Laufer, M. R., Leyland, N., Missmer, S. A., ... & Taylor, H. S. (2019). Clinical diagnosis of endometriosis: a call to action. *American journal of obstetrics and gynecology*, 220(4), 354-e1.

Apostolopoulos, N. V., Alexandraki, K. I., Gorry, A., & Coker, A. (2016). Association between chronic pelvic pain symptoms and the presence of endometriosis. *Archives of gynecology and obstetrics*, 293(2), 439-445.

Armer, J. M., Radina, M. E., Porock, D., & Culbertson, S. D. (2003). Predicting breast cancer-related lymphedema\ using self-reported symptoms. *Nursing research*, 52(6), 370-379.

Ashrafi, M., Sadatmahalleh, S. J., Akhoond, M. R., & Talebi, M. (2016). Evaluation of risk factors associated with endometriosis in infertile women. *International journal of fertility & sterility,* 10(1), 11.

Breiman, L. (2001), Random Forests, *Machine Learning* 45(1), 5-32.

Bulletti, C., Coccia, M. E., Battistoni, S., & Borini, A. (2010). Endometriosis and infertility. *Journal of assisted reproduction and genetics*, 27(8), 441-447.

Carter, J., Seed, P. T., Watson, H. A., David, A. L., Sandall, J., Shennan, A. H., & Tribe, R. M. (2020). Development and validation of predictive models for QUiPP App v. 2: tool for predicting preterm birth in women with symptoms of threatened preterm labor. *Ultrasound in Obstetrics & Gynecology*, 55(3), 357-367.

Cohen, A., Burgos-Aceves, M. A., & Smith, Y. (2016). A potential role for estrogen in cigarette smoke-induced microRNA alterations and lung cancer. *Translational lung cancer research*, 5(3), 322.

Chapron, C., Barakat, H., Fritel, X., Dubuisson, J. B., Bréart, G., & Fauconnier, A. (2005). Presurgical diagnosis of posterior deep infiltrating endometriosis based on a standardized questionnaire. *Human reproduction*, 20(2), 507-513.

Cramer, D. W., & Missmer, S. A. (2002). The epidemiology of endometriosis. *Annals of the New York Academy of Sciences*, 955(1), 11-22.

Droz, J., & Howard, F. M. (2011). Use of the Short-Form McGill Pain Questionnaire as a diagnostic tool in women with chronic pelvic pain. *Journal of minimally invasive gynecology*, 18(2), 211-217.

Eskenazi, B., & Warner, M. L. (1997). Epidemiology of endometriosis. *Obstetrics and gynecology clinics of North America*, 24(2), 235-258.

Eskenazi, B., Warner, M., Bonsignore, L., Olive, D., Samuels, S., & Vercellini, P. (2001). Validation study of nonsurgical diagnosis of endometriosis. *Fertility and sterility*, 76(5), 929-935.

Fedele, L., Bianchi, S., Carmignani, L., Berlanda, N., Fontana, E., & Frontino, G. (2007). Evaluation of a new questionnaire for the presurgical diagnosis of bladder endometriosis. *Human reproduction,* 22(10), 2698-2701.

Fuldeore, M. J., & Soliman, A. M. (2017). Prevalence and symptomatic burden of diagnosed endometriosis in the United States: national estimates from a cross-sectional survey of 59,411 women. *Gynecologic and obstetric investigation*, 82(5), 453-461.

Garavaglia, E., Ricci, E., Cioffi, R., Trojano, G., Bulfoni, A., Parazzini, F., & Viganò, P. (2017). Smoking habits and endometriosis risk among infertile women: *Results from a case control study. Italian Journal of Gynaecology and Obstetrics*, 29(3), 15-18.

Giudice, L. C. (2010). Endometriosis. *New England Journal of Medicine*, 362(25), 2389-2398.

Griffiths, A. N., Koutsouridou, R. N., & Penketh, R. J. (2007). Predicting the presence of rectovaginal endometriosis from the clinical history: a retrospective observational study. *Journal of Obstetrics and Gynaecology,* 27(5), 493-495.

Heitmann, R. J., Langan, K. L., Huang, R. R., Chow, G. E., & Burney, R. O. (2014). Premenstrual spotting of ≥ 2 days is strongly associated with histologically confirmed endometriosis in women with infertility. *American journal of obstetrics and gynecology*, 211(4), 358-e1.

Hemmert, R., Schliep, K. C., Willis, S., Peterson, C. M., Louis, G. B., Allen‐Brady, K., ... & Smith, K. R. (2019). Modifiable life style factors and risk for incident endometriosis. *Paediatric and perinatal epidemiology*, 33(1), 19-25.

Hudelist, G., Fritzer, N., Thomas, A., Niehues, C., Oppelt, P., Haas, D., ... & Salzer, H. (2012). Diagnostic delay for endometriosis in Austria and Germany: causes and possible consequences. *Human reproduction*, 27(12), 3412-3416.

Kennedy, S., Bergqvist, A., Chapron, C., D'Hooghe, T., Dunselman, G., Greb, R., ... & Saridogan, E. (2005). ESHRE guideline for the diagnosis and treatment of endometriosis. *Human reproduction,* 20(10), 2698-2704.

Lafay Pillet, M. C., Huchon, C., Santulli, P., Borghese, B., Chapron, C., & Fauconnier, A. (2014). A clinical score can predict associated deep infiltrating endometriosis before surgery for an endometrioma. *Human reproduction*, 29(8), 1666-1676.

Matalliotakis, I. M., Arici, A., Cakmak, H., Goumenou, A. G., Koumantakis, G., & Mahutte, N. G. (2008). Familial aggregation of endometriosis in the Yale Series. *Archives of gynecology and obstetrics*, 278(6), 507-511.

Mcleod, B. S., & Retzloff, M. G. (2010). Epidemiology of Endometriosis:: An Assessment of Risk Factors. *Clinical obstetrics and gynecology,* 53(2), 389-396.

Meuleman, C., Vandenabeele, B., Fieuws, S., Spiessens, C., Timmerman, D., & D'Hooghe, T. (2009). High prevalence of endometriosis in infertile women with normal ovulation and normospermic partners. *Fertility and sterility*, 92(1), 68-74.

Missmer, S. A., Hankinson, S. E., Spiegelman, D., Barbieri, R. L., Marshall, L. M., & Hunter, D. J. (2004). Incidence of laparoscopically confirmed endometriosis by demographic, anthropometric, and lifestyle factors. *American journal of epidemiology*, 160(8), 784-796.

Nnoaham, K. E., Hummelshoj, L., Kennedy, S. H., Jenkinson, C., Zondervan, K. T., & World Endometriosis Research Foundation Women's Health Symptom Survey Consortium. (2012). Developing symptom-based predictive models of endometriosis as a clinical screening tool: results from a multicenter study. *Fertility and sterility,* 98(3), 692-701.

Nnoaham, K. E., Sivananthan, S., Hummelshoj, L., Jenkinson, C., Webster, P., Kennedy, S. H., & Zondervan, K. T. (2009). Multi-center studies of the global impact of endometriosis and the predictive value of associated symptoms. *Journal of endometriosis*, 1(1), 36-45.

Paulson, J. D. (2009). Correlation of anterior vaginal wall pain with endometriosis in infertile patients. *The Journal of reproductive medicine,* 54(3), 145.

Perelló, M., Martínez-Zamora, M. A., Torres, X., Munrós, J., Llecha, S., De Lazzari, E., ... & Carmona, F. (2017). Markers of deep infiltrating endometriosis in patients with ovarian endometrioma: a predictive model. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 209, 55-60.

Peterson, C. M., Johnstone, E. B., Hammoud, A. O., Stanford, J. B., Varner, M. W., Kennedy, A., ... & Louis, G. M. B. (2013). Risk factors associated with endometriosis: importance of study population for characterizing disease in the ENDO Study. *American journal of obstetrics and gynecology*, 208(6), 451-e1.

Roland, L. T., Gurrola, J. G., Loftus, P. A., Cheung, S. W., & Chang, J. L. (2020, May). Smell and taste symptom‐based predictive model for COVID‐19 diagnosis. *In International Forum of Allergy & Rhinology*.

Schliep, K. C., Mumford, S. L., Peterson, C. M., Chen, Z., Johnstone, E. B., Sharp, H. T., ... & Buck Louis, G. M. (2015). Pain typology and incident endometriosis. *Human reproduction,* 30(10), 2427-2438.

Soliman, A. M., Taylor, H., Bonafede, M., Nelson, J. K., & Castelli-Haley, J. (2017). Incremental direct and indirect cost burden attributed to endometriosis surgeries in the United States. *Fertility and sterility,* 107(5), 1181-1190.

Staal, A. H. J., Van Der Zanden, M., & Nap, A. W. (2016). Diagnostic delay of endometriosis in the Netherlands. *Gynecologic and obstetric investigation*, 81(4), 321-324.

Surrey, E., Carter, C. M., Soliman, A. M., Khan, S., DiBenedetti, D. B., & Snabes, M. C. (2017). Patient-completed or symptom-based screening tools for endometriosis: a scoping review. *Archives of gynecology and obstetrics,* 296(2), 153-165.

Tanbo, T., & Fedorcsak, P. (2017). Endometriosis‐associated infertility: aspects of pathophysiological mechanisms and treatment options. *Acta obstetricia et gynecologica Scandinavica*, 96(6), 659-667.