

---

Abstract

# Predicting Obstetric Disease With Machine Learning Applied to Patient-Reported Data

---

Danielle Bradley, MPH, MS; Erin Landau; Adam Wolfberg, MD, MPH; Alex Baron, PhD

Ovia Health, Boston, MA, United States

---

**Corresponding Author:**

Danielle Bradley, MPH, MS

Ovia Health

76 Summer Street

Boston, MA,

United States

Phone: 771 1711

Email: [dani.bradley@oviahealth.com](mailto:dani.bradley@oviahealth.com)

---

## Abstract

---

**Background:** The rise of highly engaging digital health mobile apps over the past few years has created repositories containing billions of patient-reported data points that have the potential to inform clinical research and advance medicine.

**Objective:** To determine if self-reported data could be leveraged to create machine learning algorithms to predict the presence of, or risk for, obstetric outcomes and related conditions.

**Methods:** More than 10 million women have downloaded Ovia Health's three mobile apps (Ovia Fertility, Ovia Pregnancy, and Ovia Parenting). Data points logged by app users can include information about menstrual cycle, health history, current health status, nutrition habits, exercise activity, symptoms, or moods. Machine learning algorithms were developed using supervised machine learning methodologies, specifically, Gradient Boosting Decision Tree algorithms. Each algorithm was developed and trained using anywhere from 385 to 5770 features and data from 77,621 to 121,740 app users.

**Results:** Algorithms were created to detect the risk of developing preeclampsia, gestational diabetes, and preterm delivery, as well as to identify the presence of existing preeclampsia. The positive predictive value (PPV) was set to 0.75 for all of the models, as this was the threshold where the researchers felt a clinical response—additional screening or testing—would be reasonable, due to the likelihood of a positive outcome. Sensitivity ranged from 24% to 75% across all models. When PPV was adjusted from 0.75 to 0.52, the sensitivity of the preeclampsia prediction algorithm rose from 24% to 85%. When PPV was adjusted from 0.75 to 0.65, the sensitivity of the preeclampsia detection or diagnostic algorithm increased from 37% to 79%.

**Conclusions:** Algorithms based on patient-reported data can predict serious obstetric conditions with accuracy levels sufficient to guide clinical screening by health care providers and health plans. Further research is needed to determine whether such an approach can improve outcomes for at-risk patients and reduce the cost of screening those not at risk. Presenting the results of these models to patients themselves could also provide important insight into otherwise unknown health risks.

(*iproc* 2018;4(2):e11766) doi: [10.2196/11766](https://doi.org/10.2196/11766)

---

**KEYWORDS**

machine learning; mobile health; positive predictive value; obstetrics

---

*Edited by T Hale; this is a non-peer-reviewed article. Submitted 01.08.18; accepted 29.08.18; published 17.09.18.*

*Please cite as:*

Bradley D, Landau E, Wolfberg A, Baron A

Predicting Obstetric Disease With Machine Learning Applied to Patient-Reported Data

*iproc* 2018;4(2):e11766

URL: <http://www.iproc.org/2018/2/e11766/>

doi: [10.2196/11766](https://doi.org/10.2196/11766)

PMID:

©Danielle Bradley, Erin Landau, Adam Wolfberg, Alex Baron. Originally published in Iproceedings (<http://www.iproc.org>), 17.09.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in Iproceedings, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.iproc.org/>, as well as this copyright and license information must be included.