The application of artificial intelligence in reproductive medicine: baby steps

In recent years, with advances in big data technology, artificial intelligence (AI) has been used in multiple fields, including health care. Assisted reproductive technology (ART) includes many interrelated steps and requires many linked clinical decisions. In the field of ART, there has been increasing interest in the use of AI for aiding diagnosis, tailoring treatment, and predicting prognosis (1). The application of AI in ART has been investigated in nearly all aspects of patient care, including sperm evaluation, embryo selection, and the optimization of ovarian stimulation (1).

Ovarian stimulation is an integral step of ART, with the goal of retrieving multiple mature oocytes to increase the chances of pregnancy. A trigger shot with human chorionic gonadotropin and/or a gonadotropin-releasing hormone agonist is administered in the late follicular phase to induce final oocyte maturation. The timing of trigger is correlated with oocyte developmental competence and is one of the critical decisions made by clinicians. However, the decision on the timing of trigger is usually based on the clinicians' experience of incorporating multiple factors, such as the size of leading follicles, hormone levels, the duration of stimulation, the results of previous stimulation, and the risks of ovarian hyperstimulation syndrome and premature ovulation. The criteria for the timing of trigger vary greatly among different centers and clinicians (2). The optimal timing of trigger lacks evidence-based consensus and lends itself to AI solutions.

In the article by Loewke (3), published in the current issue of *Fertility and Sterility*, sets of machine learning models were developed to predict the optimal day of trigger during ovarian stimulation. Based on follicle count, size, and estradiol level, linear models were used to predict the number of metaphase II (MII) oocytes when triggered on the same day compared with those triggered the next day. When the predicted number of MII oocytes triggered the next day is higher than that of MII oocytes triggered on the same day, the continuation of stimulation is recommended, otherwise triggering on the same day is recommended. The result of the analysis of retrospective data showed that the use of the models to optimize the day of trigger could increase the yield by 2–3 MII oocytes.

The increment is of clinical significance because the number of oocytes in other studies has been positively correlated with the rate of cumulative live birth rate (4), although the investigators of this article were unable to obtain accurate information on live birth outcomes from their dataset. This is clearly an opportunity lost because mature oocytes and embryo development are just surrogate outcomes for the desired goal of live births (as the investigators acknowledged). Further, another aspect lacking from the study is the incorporation of adverse event rates, specifically mild, moderate, or severe ovarian hyperstimulation syndrome rates. Every clinical decision is a risk-benefit analysis, and it is tempting to speculate that the predominant category of early triggers by the model (48.7% as opposed to 37.5% on-time triggers and 13.8% late triggers) was guided as much by the avoidance of hyperstimulation as by the optimal yield of oocytes. We have long advocated that infertility studies should focus on patient-relevant clinical outcomes, with full assessment of all adverse outcomes in the mother, father, and infant in these studies (5).

Among variables included in the models, it was found that follicles with a size of 14–15 mm and 16–17 mm were more closely correlated with the predicted number of MII oocytes than follicles with a size greater than 19 mm. This was an interesting finding because the most frequently used criterion for the timing of trigger was several leading follicles between the size of 16 and 22 mm (2). Although not confirmed, the finding suggested that follicles with a size of 14–17 mm are more likely to yield MII oocytes than follicles of a larger size. This result may also, at least partially, explain that nearly 50% of cycles were identified as early trigger when the actual day of trigger was compared with the model's recommendation. Further studies are warranted to evaluate whether we should pay more attention to medium-sized follicles for determining the timing of trigger when AI assistance is unavailable.

Another important characteristic of their model is interpretability compared with the common black-box algorithm in machine learning models. As the investigators stated, when the prediction performance is comparable with that of other similar studies, model interpretability may facilitate clinicians' understanding and acceptance of the models. The other advantages of interpretable models may include the potential to reveal a hidden relationship and the possibility to overcome ethical concerns. Thus, at this beginning stage of the era of AI in ART, the interpretable models may more easily progress to clinical application compared with the black-box algorithm.

This is an intriguing "baby" step toward the application of AI in improving patients' outcomes after ART. Obviously, replication using other independent, retrospective datasets would prepare the playing field for prospective randomized trials of humans vs. machines in determining the timing of trigger. These should be based on full accounting of pregnancy outcomes and adverse events. If it is anything like chess battles between man and machine, prepare for the demise of man—with apologies to Gary Kasparov—and a shout out to International Business Machine's Deep Blue. However, chess is just a game, and this is a matter of life; so let us see how this plays out.

Daimin Wei, M.D.[a,b]
Richard S. Legro, M.D.[c]
Zi-Jiang Chen, M.D.[a,b]

[a] Center for Reproductive Medicine, Shandong University, Jinan, People's Republic of China; [b] Medical Integration and Practice Center, Shandong University, Jinan, People's Republic of China; and [c] Department of Obstetrics and Gynecology, Penn State College of Medicine, Hershey, Pennsylvania

https://doi.org/10.1016/j.fertnstert.2022.05.002

DIALOG: you can discuss this article with its authors and other readers at https://www.fertstertdialog.com/posts/35213

REFERENCES

1. Zaninovic N, Rosenwaks Z. Artificial intelligence in human in vitro fertilization and embryology. Fertil Steril 2020;114:914–20.
2. ESHRE Guideline Group on Ovarian StimulationBosch E, Broer S, et al. ESHRE guideline: ovarian stimulation for IVF/ICSI. Hum Reprod Open 2020;2020: hoaa009.
3. Loewke K. An interpretable machine learning model for predicting the optimal day of trigger during ovarian stimulation. Fertil Steril 2022;118: 101–8.
4. Polyzos NP, Drakopoulos P, Parra J, Pellicer A, Santos-Ribeiro S, Tournaye H, et al. Cumulative live birth rates according to the number of oocytes retrieved after the first ovarian stimulation for in vitro fertilization/intracytoplasmic sperm injection: a multicenter multinational analysis including ~15,000 women. Fertil Steril 2018;110:661–70.
5. Harbin Consensus Conference Workshop Group. Improving the reporting of clinical trials of infertility treatments (IMPRINT): modifying the CONSORT statement. Fertil Steril 2014;102:952–9.