Artificial intelligence in the service of intrauterine insemination and timed intercourse in spontaneous cycles

Michal Youngster, M.D.,[a,b] Almog Luz, B.Sc.,[c] Micha Baum, M.D.,[b,c,d,e] Rohi Hourvitz, B.Sc.,[c]
Shachar Reuvenny, M.Sc.,[c] Ettie Maman, M.D.,[b,c,d,e] and Ariel Hourvitz, M.D., M.H.A.[a,b,c]

[a] IVF Unit, Department of Obstetrics and Gynecology, Shamir Medical Center, Zerifin, Israel; [b] Sackler School of Medicine, Tel Aviv University, Tel-Aviv, Israel; [c] FertilAi, Ramat Gan, Israel; [d] IVF Unit, Herzliya Medical Centre, Herzliya, Israel; and [e] IVF Unit, Department of Obstetrics and Gynecology, Sheba Medical Center, Ramat-Gan, Israel

Objective: To develop a machine learning model designed to predict the time of ovulation and optimal fertilization window for performing intrauterine insemination or timed intercourse (TI) in natural cycles.
Design: A retrospective cohort study.
Setting: A large in vitro fertilization unit.
Patient(s): Patients who underwent 2,467 natural cycle–frozen embryo transfer cycles between 2018 and 2022.
Intervention(s): None.
Main Outcome Measure(s): Prediction accuracy of the optimal day for performing insemination or TI.
Result(s): The data set was split into a training set including 1,864 cycles and 2 test sets. In the test sets, ovulation was determined according to either expert opinion, with 2 independent fertility experts determining ovulation day (“expert”) (496 cycles), or according to the disappearance of the leading follicle between 2 consecutive days’ ultrasound examinations (“certain ovulation”) (107 cycles). Two algorithms were trained: an NGBoost machine learning model estimating the probability of ovulation occurring on each cycle day and a treatment management algorithm using the learning model to determine an optimal insemination day or whether another blood test should be performed. The estradiol progesterone and luteinizing hormone levels on the last test performed were the most influential features used by the model. The mean numbers of tests were 2.78 and 2.85 for the “certain ovulation” and “expert” test sets, respectively. In the “expert” set, the algorithm correctly predicted ovulation and suggested day 1 or 2 for performing insemination in 92.9% of the cases. In 2.9%, the algorithm predicted a “miss,” meaning that the last test day was already ovulation day or beyond, suggesting avoiding performing insemination. In 4.2%, the algorithm predicted an “error,” suggesting performing insemination when in fact it would have been performed on a nonoptimal day (0 or −3). The “certain ovulation” set had similar results.
Conclusion(s): To our knowledge, this is the first study to implement a machine learning model, on the basis of the blood tests only, for scheduling insemination or TI with high accuracy, attributed to the capability of the algorithm to integrate multiple factors and not rely solely on the luteinizing hormone surge. Introducing the capabilities of the model may improve the accuracy and efficiency of ovulation prediction and increase the chance of conception.
Clinical Trial Registration Number: HMC-0008-21. (Fertil Steril® 2023;120:1004–12. ©2023 by American Society for Reproductive Medicine.)
El resumen está disponible en Español al final del artículo.

Key Words: Artificial intelligence, IUI, timed intercourse, ovulation, natural cycle

Predicting ovulation is an important unmet need in fertility care. Although accurate prediction of ovulation can benefit several populations trying to conceive, including patients undergoing intrauterine insemination (IUI) on the basis of a natural cycle, self-intracervical insemination (ICI), and timed intercourse (TI), simple reliable methods that will serve patients and caregivers are lacking.

The demand for IUI has increased in recent years, especially with the use of donor sperm (1). The first-line treatment for young women using donor sperm for conception, as well as for couples with isolated mildly reduced sperm quality, is the spontaneous cycle combined with IUI to avoid multiple pregnancy (2, 3). The time of ovulation following a naturally occurring luteinizing hormone (LH) surge is subject to higher variability than following human chorionic gonadotropin (hCG) administration. Available data show

Received February 4, 2023; revised July 6, 2023; accepted July 18, 2023; published online July 23, 2023.
E.M. and A.H. should be considered similar in author order.
Supported by FertilAI, Israel.
Presented at the 78th Annual Meeting of the American Society for Reproductive Medicine, Anaheim, California, October 22–26, 2022.

Correspondence: Michal Youngster, M.D., IVF Unit, Department of Obstetrics and Gynecology, Shamir Medical Center, Zerifin 70300, Israel (E-mail: michalyo@gmail.com).

Fertility and Sterility® Vol. 120, No. 5, November 2023 0015-0282/$36.00

https://doi.org/10.1016/j.fertnstert.2023.07.008

that ovulation can occur within a wide time frame, ranging from 22 to 56 hours after the onset of the LH surge (4–8). This can be explained by the lack of consensus as to the definition of the LH surge and the methods used for its detection (9, 10). The variable timing of ovulation, combined with the need to closely monitor the levels to detect the spontaneous LH surge, makes correct scheduling of IUI based solely on the detection of the spontaneous LH surge a challenging procedure. On the basis of previous studies, the optimal timing for IUI in natural cycles is thought to be 1 day after the LH surge (3, 8).

The window of opportunity in timing of intercourse or self-ICI is wider and more flexible because of the possibility of repeated exposure. The fertile window, in which a women may become pregnant, lasts approximately 6 days and may vary considerably even in women with regular cycles (11). Therefore, a Cochrane review concluded that couples using TI have higher chances of conceiving (12). The chances for conception reach 30% 1–2 days before ovulation, 7% on the day of ovulation, and only 1.5% on the day after ovulation (13, 14).

Several accepted methods for timing of insemination or TI are used. Available fertility awareness methods, such as ovulation detection kits on the basis of urine LH and estrogen and cervical mucus monitoring, have been shown to increase the probability of conceiving but are considered less accurate, present false-positive and false-negative results (15), and may not be suitable for all couples. Another more resource-intensive method requires follow-up by a fertility expert with repeated ultrasound (US) examinations combined with the blood levels of LH.

In recent years, the artificial intelligence (AI) techniques have emerged as objective standardized methods to improve the outcome of fertility treatments, mostly in personalized ovarian stimulation (16–19). To our knowledge, no study has developed an AI model to assist scheduling the best time for IUI, ICI, or intercourse on the basis of ovulation prediction. The lack of a reliable scheduling method, combined with several populations in demand for such, triggered our current study. We have implemented a machine learning technology, using natural cycle–frozen embryo transfer (NC-FET) cycles with the possibility of retrospectively determining accurate ovulation timing, to be able to determine the optimal fertile window for each method of conception. This study aimed to show whether this technology can predict optimal timing for IUI, self-ICI, or intercourse in spontaneous cycles, based solely on blood test results without the need for sonographic data, providing a novel tool to assist caregivers and patients.

MATERIALS AND METHODS

This is a retrospective cohort study conducted at Herzliya Medical Center, Israel. The study was approved by the institutional review board (HMC-0008-21). The data set consisted of all NC-FET cycles with adequate documentation, performed between September 2018 and July 2021. Each cycle included information about the patients' age at the time of transfer; body mass index (BMI); hormonal levels including estradiol, progesterone, and LH from at least 2 blood tests; and size of the follicles from at least 2 US scans.

Data Preparation

Filtering The data set was filtered to include only cycles in which the patient was ovulating between days 7 and 21, resulting in a total of 2,467 of 2,719 cycles.

Missing Values Blood tests without all 3 hormone levels were removed. The missing values of weight and height were handled by the NGBoost's sparsity aware split finding algorithm, which determines the optimal value according to its training data (20).

Augmentation Several dozen features, comprising basic characteristics and computed values that may improve the accuracy of the model because of the biologic or model-related mathematical reasons were created after discussion between expert physicians and data scientists. An automatic tuning process determined which features were beneficial to the model by repeatedly testing its performance using different combinations. The combination of 16 final features resulted in the optimal performance on a holdback validation set: ovulation day; blood tests' cycle days; patients' age, weight, height, and BMI; and 3 features per hormone (LH, estradiol, and progesterone)—its absolute value, difference from the previous test, and deviation from the previous test (the difference divided by the number of days between tests).

Train/test split A total of 107 cycles in the data set had 2 consecutive US examinations demonstrating the disappearance of a leading follicle. These cycles were set aside as a "certain ovulation" test set, resulting in 2,360 remaining cycles.

The remaining cycles were randomly split into a training set with 1,864 cycles and an "expert" test set with 496 cycles, resulting in an approximate 76/20/4 split with 76% of the data in the training set, 20% in the random test set, and 4% in the "certain ovulation" test set.

To train and evaluate the algorithms, it was necessary to determine the ovulation day for each cycle in the data set. For the training set, a learning approach called student-teacher was employed. This approach entailed a previously trained, and accurate, ovulation prediction algorithm (the "teacher") that was used to predict the ovulation day in each cycle using all available cycle data, whereas the new algorithm (the "student") used these predictions as a target in its training. The teacher algorithm in this case was a FET ovulation prediction algorithm that used all available cycle data, including both blood and US tests before and after ovulation, to determine the ovulation day as accurately as possible. On the other hand, the student algorithm was trained using only 1–2 blood tests from each cycle, without sonographic data.

To eliminate any bias, a different method was used to determine the day of ovulation in the 2 test sets. In the "expert" test set, 2 fertility experts reviewed each chart and independently determined the ovulation day. The final result was then determined by majority decision between the 3 physicians (the 2 experts who retrospectively reviewed the data and the original decision of the physician performing the transfer). There were no instances without majority agreement. In the "certain ovulation" test set, ovulation timing was determined according to the disappearance of the leading follicle between 2 consecutive days of US examinations.

Algorithm Design

We aimed to determine whether a machine learning algorithm can effectively manage the IUI treatment process, relying on the blood test results only. The management included suggesting the optimal timing for the next blood test, until a reliable prediction of ovulation can be made and providing a reliable suggestion for the optimal insemination day. The algorithms' suggestion was considered "correct" if it was given in advance and suggested a day exactly 1 or 2 days before ovulation (day −1 or −2). A "missed" ovulation was defined when the algorithm predicted the latest blood test at the day of ovulation or later (day 0, +1, or +2), meaning that the algorithm predicted that it may be too late for performing IUI. An "error" was defined when the algorithm suggested an insemination day at or after ovulation or >2 days before ovulation, meaning that insemination would not be performed within the optimal defined insemination window.

For the purpose of the study, 2 algorithms were developed: an ovulation prediction model and a treatment management algorithm. A detailed explanation of the algorithms is shown in the Supplemental Materials (available online)—materials and methods—indicating the following, in brief:

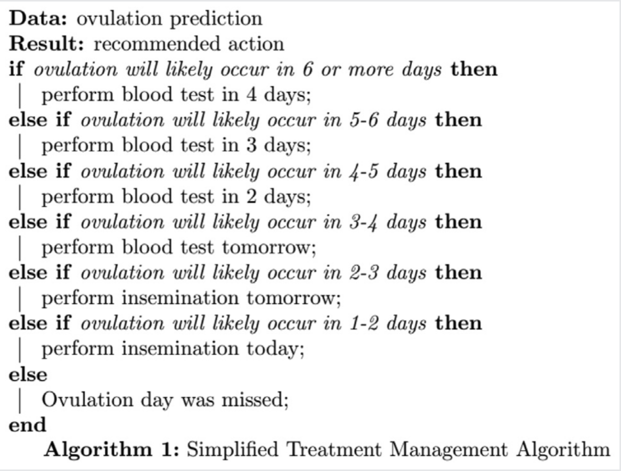
- An ovulation prediction model: An NGBoost machine learning model that estimated the probability of ovulation occurring on each cycle day on the basis of the available visit's data.
- A treatment management algorithm that used the learning model after receiving a new visit's data to determine whether an optimal insemination day can be determined or another blood test should be performed, in which case the optimal day to perform the additional test is provided. The goal was to predict ovulation with a minimum number of tests while avoiding missing the ovulation day. A simplified treatment management algorithm is presented in Figure 1.

Confidence intervals (CIs) were calculated using the Wilson score and Monte Carlo method. Multiclass calibration using 1 vs. rest methodology was performed as a part of the models' development.

RESULTS

The mean age of participants in the data set was 36.3 ± 5.8 years. The mean BMI was 23.8 ± 5.0. The mean number of past pregnancies was 0.7, with the past numbers of retrievals and transfers of 1.3 and 1.6, respectively.

FIGURE 1



```
Data: ovulation prediction
Result: recommended action
if ovulation will likely occur in 6 or more days then
    perform blood test in 4 days;
else if ovulation will likely occur in 5-6 days then
    perform blood test in 3 days;
else if ovulation will likely occur in 4-5 days then
    perform blood test in 2 days;
else if ovulation will likely occur in 3-4 days then
    perform blood test tomorrow;
else if ovulation will likely occur in 2-3 days then
    perform insemination tomorrow;
else if ovulation will likely occur in 1-2 days then
    perform insemination today;
else
    Ovulation day was missed;
end
    Algorithm 1: Simplified Treatment Management Algorithm
```

A simplified treatment management algorithm logic.

Youngster. AI for IUI and timed intercourse in NC. Fertil Steril 2023.

Ovulation Prediction Model

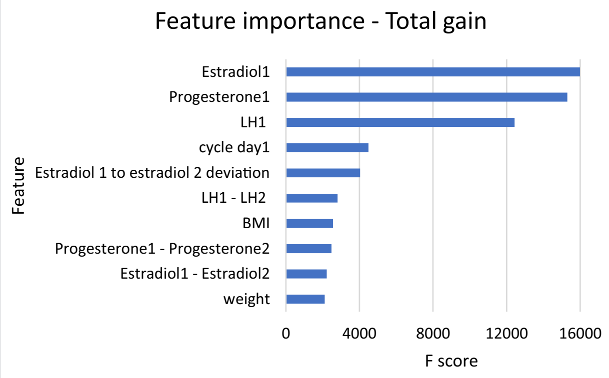
The 10 most influential features used by the ovulation prediction model to accurately determine ovulation day are presented in Figure 2. The estradiol and progesterone levels on the last test day were the most influential parameters, followed by the LH levels.

The ovulation prediction model results are presented separately for the "expert" and "certain ovulation" test sets. They can best be displayed as a series of confusion matrices (Supplemental Materials and Supplemental Fig. 2A and B).

Treatment Management Algorithm

Treatment management algorithm performance is presented separately for the "expert" and "certain ovulation" test sets. The model performed best when calling in the patient for

FIGURE 2



Feature importance. The number 1 indicates the last test, whereas 2 indicates the previous test. BMI = body mass index; LH = luteinizing hormone.

Youngster. AI for IUI and timed intercourse in NC. Fertil Steril 2023.

the first blood test on day 8 of her menstrual cycle (the cycle length from previous cycles of the same patient was not available). Table 1 presents the performance of algorithms for each possible ovulation day for both test sets, given that the patient undergoes the initial blood test on day 8 of her cycle.

“Expert” test set The mean number of tests for all instances included was 2.85 (95% CI, 2.83–2.87) (Table 1). The longer the cycle, the more tests needed before an ovulation could be accurately predicted. If a patient arrived for her first test on day 8 and ovulated on day 7 or 8, the success rates were 0% because ovulation had already occurred (probability rates for ovulation, 0.4% and 0.7%, respectively). The highest ovulation probability was on days 13–15 of the menstrual cycle (12.3%–12.4%). Ovulation prediction was successful in 91%–97% for most of the cycles.

“Certain ovulation” test set The mean number of tests for all instances included was 2.78 (95% CI, 2.73–2.82) (Table 1). The highest ovulation probability was on days 13–15 of the menstrual cycle (12.3%–12.4%). Ovulation prediction was successful in 88%–99% for most of the cycles.

The probability distribution of our algorithmic predictions and distributions of the true values for each predicted value for both test sets are presented in Table 2.

“Expert” test set The algorithm could correctly predict ovulation and suggest day −1 or −2 for performing IUI in 92.9% of the cases (95% CI, 92.3–93.6) (Table 2). In approximately 70% of the cases, the last blood test predicted the day to be −2 or −3, suggesting an IUI next day. Of these, 57% were actually −2, meaning that IUI was scheduled the next day at −1. Approximately 41% were −3, meaning that IUI was performed on day −2. In approximately a quarter of all cases, the prediction of the last blood test was day −2 or −1, with IUI suggested the same day, because approximately half were, indeed, day −1. In 2.9% of the cases (95% CI, 2.6–3.2), the algorithm predicted the last test day to be ovulation day or beyond. These were considered “missed,” with a suggestion to avoid performing IUI in that cycle. In 4.2% of the cycles (95% CI, 3.6–4.8), the algorithm predicted the last test day to be −2 or −3 with a suggestion to perform IUI the next day, when in fact it was a different day and IUI would have been performed on a

TABLE 1

The algorithm performance for each possible ovulation day: “expert” and “certain ovulation” test sets.

Ovulation day	Ovulation probability	Success rate	Mean tests
“Expert” test set			
7	0.4%	0%	1.01
8	0.7%	0%	1.01
9	1.8%	95%	1.01
10	3.3%	98%	1.07
11	5.8%	91.3%	1.52
12	9.3%	90.9%	2.16
13	12.4%	93.5%	2.43
14	12.4%	97.1%	2.37
15	12.3%	94.9%	2.74
16	10.7%	97%	3.36
17	9.8%	95.4%	3.53
18	7.1%	91.9%	3.68
19	5.8%	91.8%	4.07
20	5.2%	91.1%	4.41
21	2.9%	89.6%	4.60
“Certain ovulation” test set			
Ovulation day	**Ovulation probability**	**Success rate**	**Mean tests**
7	0.4%	0%	1.02
8	0.7%	0%	1.02
9	1.8%	94.8%	1.00
10	3.3%	100%	1.00
11	5.8%	89.8%	1.57
12	9.3%	84.4%	1.96
13	12.4%	97.7%	2.62
14	12.4%	98.4%	2.20
15	12.3%	92.8%	2.76
16	10.7%	89.6%	3.22
17	9.8%	91.8%	3.59
18	7.1%	98. 4%	3.35
19	5.8%	98.7%	3.77
20	5.2%	88.0%	4.21
21	2.9%	94.1%	4.64

Note: The first test performed on day 8 of the menstrual cycle.

Youngster. AI for IUI and timed intercourse in NC. Fertil Steril 2023.

TABLE 2

Probability distribution of algorithmic predictions and distributions of the true values for each predicted value: "expert" and "certain ovulation" test sets.

Expert		**True value**						
Predicted class	**Probability**	**≤−6**	**−5**	**−4**	**−3**	**−2**	**−1**	**≥0**
−3:−2	69.6%	0.9%	0.2%	1%	40.6%	56.9%	0.4%	0%
−2:−1	24.8%	0.9%	0.6%	0.3%	1.8%	47.0%	49.2%	0.1%
−1	2.8%	16.3%	0%	0%	29.8%	6.3%	44.3%	3.2%
≥0	2.9%	0%	5.1%	2.9%	1.1%	17.1%	19.3%	54.5%

Certain ovulation		**True value**						
Predicted class	**Probability**	**≤−6**	**−5**	**−4**	**−3**	**−2**	**−1**	**≥0**
−3:−2	61.6%	0%	0%	1.9%	53.1%	44.8%	0.2%	0%
−2:−1	32.9%	0%	0%	0%	3.4%	81.1%	15%	0.5%
−1	1.4%	0%	0%	23.6%	0%	0%	62.1%	14.4%
≥0	4.1%	0%	0%	43.6%	0%	0%	7.3%	49.1%

Note: The predicated class was defined as the class predicted by the algorithm on the last test day. Probability was defined as the probability of arriving at the predicted class when following the treatment management suggestions. The true value was defined as the distribution of the actual days relative to ovulation for a given predicted day of the algorithm. The green highlight indicates success. The red highlight indicates error. The yellow highlight indicates missed,

Youngster. AI for IUI and timed intercourse in NC. Fertil Steril 2023.

nonoptimal day (0 or −3). These cases were considered "error." When IUI was to be scheduled the same day of the last test (predicted day −1 or −1:−2), the model provided an early indication for it in 71% (prediction day −1:−2) and 63% (day −1) of the cases, with a mean of 70% of the cases overall when taking into account the probability of each prediction.

"Certain ovulation" test set The success rate was similar (92.4%; 95% CI, 91.3–93.6) with a mean test count of 2.78 (95% CI, 2.73–2.82) (Table 2). The miss rate was 4.5% (95% CI, 3.8–5.1), with an error rate of 3.1% (95% CI, 2.2–4.1). When IUI was to be scheduled the same day of the last test (day −1 or −2), the model provided an early indication in 63% (day −1:−2) and 42% (day −1) of the cases, with a mean of 62% of the cases overall.

Alternative configurations The same prediction model can be applied to different management algorithms' configurations according to the institutional preferences. For example, if same-day IUI is less feasible, then more frequent blood tests may significantly reduce the chance of its occurrence *(*Supplemental Table 2*)*. Increasing the number of blood tests to a mean of 3.31 (95% CI, 3.29–3.33) resulted in a success rate of 93.6% (95% CI, 92.9–94.3), with a miss rate of 2.1% (95% CI, 1.8–2.3) and an error rate of 4.3% (95% CI, 3.7–5.0). This configuration reduced the number of same-day IUIs to only 15.6%, with >65% of the cases with early warning. Furthermore, approximately 70% of these cases were actually −2; thus, if they were predicted to be −2 with high certainty, then IUI can safely be scheduled for the following day.

DISCUSSION

This is the first study evaluating the use of AI for determining the optimal timing for IUI or intercourse. Based solely on hormonal profile, with a mean of 2.85 blood tests per cycle, the model accurately determined timing for IUI in 92.9% and 92.4% of the cases, with error rates of 4.2% and 3.1% for the "expert" and "certain ovulation" test sets, respectively.

As opposed to the NC-FET cycles in which follicle rupture can be taken into account retrospectively when scheduling the transfer, the main challenge in optimizing timing for IUI is that it should be performed 1 day before ovulation and, ideally, planned a day ahead. However, most hormonal changes, including the increases in the LH and progesterone levels and decrease in the estradiol level, take place on days −1 and −2 with some overlap and are difficult to anticipate on day −3. Therefore, most study investigators and clinicians base their decision for performing IUI on the identification of the onset of the LH surge. However, no standard definition of the onset of the LH surge exists, and the published literature includes numerous definitions, including an LH level of ≥10 to 20 IU/L (21–24), ≥180% or >2 times of the mean of the preceding LH values (5, 6). As such, there is a marked interindividual variation in the interval between the LH surge and actual follicle rupture, ranging from 22 to 56 hours, with a mean of 34 hours (4, 10, 25).

The optimal time for IUI is considered 1 day after the onset of the LH surge, most often a day before ovulation (day −1) (3, 13). Our algorithm could not reliably differentiate between days −2 and −3 because these days overlapped in terms of hormonal changes. However, the chance of conception is relatively similar on days −2 and −1 (13). When triggering ovulation, it was shown that IUI may be performed immediately after the hCG administration and up to 36 hours later with no influence on the pregnancy rates (26). Therefore, the algorithm was trained to schedule IUI on day −1 or −2.

Several methods are used for scheduling IUI on the basis of the LH surge. The combination of blood tests and US monitoring is probably the most accurate method currently in use but is costly and requires repeated US examinations and access to facilities. Urinary LH kits are cheap and accessible for monitoring but have several disadvantages. They require daily and even twice daily tests to reach high specificity and sensitivity (10, 27). There is a wide variation in the sensitivity of the urine assay and a time delay due to the prolonged urinary clearance. They may produce false-negative results in the lower ranges of LH values or show a premature LH surge without ovulation (10). The LH kits rely solely on the LH surge and disregard other variables. As a result, most studies, and consequently most clinicians, base the scheduling mainly on the onset of the LH surge, ignoring other hormonal changes such as progesterone and estradiol levels, and the follicle size. The precise determination of ovulation in the training set was made by an algorithm that was trained to determine ovulation day on the basis of variables from thousands of FET cycles, including the follicle size, estradiol and progesterone levels, and patient's age and BMI, in addition to the LH levels. The algorithm presented in the study was then trained using those predicted ovulation days to identify the days before ovulation on the basis of the hormonal levels only because US did not confer significant additional value of the days before ovulation. In the algorithm, the estradiol and progesterone levels had a significant value in the prediction in addition to the LH levels, with patients' age and BMI taken into account, presumably making the prediction more accurate than relying solely on the LH surge.

Human chorionic gonadotropin may be administered for ovulation induction for timing of IUI. This method has the advantage of simple scheduling of the procedure, avoiding repeated tests and weekend procedures. Nevertheless, it has several disadvantages. First, it requires repeated costly US examinations to avoid early triggering. Second, hCG is usually administered when the follicle is ≥17. At this stage, a spontaneous LH surge may already have begun, with some studies showing a detrimental effect on the pregnancy rates with the use of hCG for ovulation induction. This is possibly due to immaturity of the oocyte if hCG is administered too early and reduced endometrial receptivity due to the effect of the hCG on the endometrium (28–30), as opposed to the LH effect (31, 32). Implementation of our algorithm may enable avoidance of hCG triggering. However, when needed, to avoid missed opportunities similar to the case of an approaching weekend, the algorithm may warn the physician and suggest using hCG triggering for scheduling purposes.

Timed intercourse increases the chance of conceiving and may shorten time to pregnancy, with the highest-yield timing being 1 or 2 days before ovulation (11, 12), which the model can accurately predict. Currently, several methods are available for timing of intercourse or ICI, with the urine LH kit being the most accurate but demands daily urine sampling and has false-positive or false-negative results. Furthermore, the daily use of urine LH kits to detect ovulation and TI may add significant stress in the fertile and subfertile populations (33, 34). Avoiding daily testing with unexpected results to be replaced by a process with a defined schedule may relieve the added stress, which is known to have a negative impact on the pregnancy success rates (33). Several populations, such as couples with long-distance relationship or couples having physical or emotional difficulty with frequent intercourse, will derive extra benefit from using the application for TI (11). Women with irregular long cycles can avoid repeated daily unnecessary urine tests. Anovulatory patients will be informed by the algorithm to turn to their physician for consultation after several tests without detection of ovulation, whereas in the current methods, patients cannot differentiate a prolonged follicular phase from a normal cycle in which they may have missed their LH surge. For patients undergoing ICI, especially when using frozen sperm, accurate timing is important because the fertile window is likely to be shorter as the cervical mucus may become impenetrable early (35).

When designing the model, we took an extremely conservative approach with strict definitions for "missed" and "error". Day 0, which was identified by the model in 3% of the cases, was declared as "missed." In approximately half of those cases, it was indeed day 0, and in approximately 20%, it was day −1 or −2. Patients and physicians may still consider performing IUI early that day, possibly after US to present a preruptured follicle. We defined "error" in 3%–4% of the cases. In these cases, the model incorrectly identifies the day for scheduling IUI when in fact it is not a recommended day for performing the procedure, either day −3 or 0. Despite the "error" definition in our algorithm, there is still a 15%–20% chance of conceiving even on day 0 or −3 (13). These considerations may reduce the true "missed" and "error" rates and improve the performance of our model. Moreover, timing of the first blood test may be adapted in patients, according to their expected cycle length, especially those with short or prolonged cycles, thus reducing the number of blood tests required and reducing the "missed" cycles in patients with short cycles.

The main limitation of the study is the use of retrospectively collected data. To overcome this limitation, we tested our model with 2 different test sets to validate its predictions. The success, missed, and error rates for both test sets were similar, reassuring the validity of our results. A prospective randomized control study with the outcome of pregnancy and live birth rates is warranted to further validate these results and show real-life clinical benefit from the use of the algorithm.

A second limitation stems from the fact that the data set was originally collected from the NC-FET cycles to enable accurate determination of ovulation. This may represent a substantial discrepancy in populations. Nevertheless, patients who are candidates for NC-FET cycles or for NC-IUI most likely have regular cycles; thus, the comparison of the cycle characteristics is acceptable. A third limitation is the possibility of same-day IUI. Days −2 and −3 in the last blood test were identified in approximately 70% of the cases, allowing us to perform IUI the following day. Days −1 and −2 were identified in approximately 25% of the cases. Of these, half were actually day −1, and the rest were mostly −2. In some settings, this may be a disadvantage because the laboratory results may arrive too late to perform the IUI the same day. However, the algorithm was able to warn us in advance about this possibility in most cases and, thereby, allow the physician and patient to prepare ahead of time for same-day IUI. Moreover, this study aimed to be as accurate as possible with minimum blood tests; however, our final goal was to be able to adapt this model according to the requirements of each institution's staff and patients' population. Once our prediction model algorithm can accurately predict ovulation day, it is possible to make the required adaptations in the management algorithm, for example, increasing the number of blood tests to decrease the chance of same-day IUI when not feasible as presented in our results or avoiding weekend IUI.

In conclusion, this is the first study to implement a machine learning model for scheduling IUI, TI, or self-ICI with high accuracy, attributed to the capability of the algorithm to integrate multiple factors and not rely solely on the controversial LH surge. Using this model can assist the provider in a more accurate prediction of ovulation and IUI scheduling. It may also provide a useful tool for couples to identify the fertility window for repeated intercourse and for more accurate timing if necessary and for accurate self-ICI. Further prospective studies are needed to validate our results.

Declaration of interests: M.Y. has nothing to disclose. A.L. is a shareholder and board member of FertilAI LTD. M.B. is a shareholder and board member of FertilAI LTD. R.H. is a shareholder and board member of FertilAI LTD. S.R. is an employee of FertilAI LTD. E.M. is a shareholder and board member of FertilAI LTD. A.H. is a shareholder and board member of FertilAI LTD.

REFERENCES

1. Arocho R, Lozano EB, Halpern CT. Estimates of donated sperm use in the United States: National Survey of Family Growth 1995-2017. Fertil Steril 2019;112:718–23.
2. Kop PA, Mochtar MH, O'Brien PA, Van der Veen F, van Wely M. Intrauterine insemination versus intracervical insemination in donor sperm treatment. Cochrane Database Syst Rev 2018;1:CD000317.
3. Cohlen B, Bijkerk A, Van Der Poel S, Ombelet W. IUI: review and systematic assessment of the evidence that supports global recommendations. Hum Reprod Update 2018;24:300–19.
4. Temporal relationships between ovulation and defined changes in the concentration of plasma estradiol-17 beta, luteinizing hormone, follicle-stimulating hormone, and progesterone. I. Probit analysis. World Health Organization, Task Force on Methods for the Determination of the Fertile Period, Special Programme of Research, Development and Research Training in Human Reproduction. Am J Obstet Gynecol 1980;138:383–90.
5. Testart J, Frydman R, Feinstein MC, Thebault A, Roger M, Scholler R. Interpretation of plasma luteinizing hormone assay for the collection of mature oocytes from women: definition of a luteinizing hormone surge-initiating rise. Fertil Steril 1981;36:50–4.
6. Wetzels LC, Hoogland HJ. Relation between ultrasonographic evidence of ovulation and hormonal parameters: luteinizing hormone surge and initial progesterone rise. Fertil Steril 1982;37:336–41.
7. Garcia JE, Jones GS, Wright GL Jr. Prediction of the time of ovulation. Fertil Steril 1981;36:308–15.
8. Blockeel C, Knez J, Polyzos NP, De Vos M, Camus M, Tournaye H. Should an intrauterine insemination with donor semen be performed 1 or 2 days after the spontaneous LH rise? A prospective RCT. Hum Reprod 2014;29:697–703.
9. Fuh KW, Wang X, Tai A, Wong I, Norman RJ. Intrauterine insemination: effect of the temporal relationship between the luteinizing hormone surge, human chorionic gonadotrophin administration and insemination on pregnancy rates. Hum Reprod 1997;12:2162–6.
10. Erden M, Mumusoglu S, Polat M, Yarali Ozbek I, Esteves SC, Humaidan P, et al. The LH surge and ovulation re-visited: a systematic review and meta-analysis and implications for true natural cycle frozen thawed embryo transfer. Hum Reprod Update 2022;28:717–32.
11. Practice Committee of the American Society for Reproductive Medicine and the Practice Committee of the Society for Reproductive Endocrinology and Infertility. Optimizing natural fertility: a committee opinion. Fertil Steril 2022;117:53–63.
12. Manders M, Mclindon L, Schulze B, Beckmann MM, Kremer JA, Farquhar C. Timed intercourse for couples trying to conceive. Cochrane Database Syst Rev 2015:CD011345.
13. Wilcox AJ, Weinberg CR, Baird DD. Timing of sexual intercourse in relation to ovulation. Effects on the probability of conception, survival of the pregnancy, and sex of the baby. N Engl J Med 1995;333:1517–21.
14. Stanford JB, White GL, Hatasaka H. Timing intercourse to achieve pregnancy: current evidence. Obstet Gynecol 2002;100:1333–41.
15. Leiva RA, Bouchard TP, Abdullah SH, Ecochard R. Urinary luteinizing hormone tests: which concentration threshold best predicts ovulation? Front Public Health 2017;5:320.
16. Fanton M, Nutting V, Solano F, Maeder-York P, Hariton E, Barash O, et al. An interpretable machine learning model for predicting the optimal day of trigger during ovarian stimulation. Fertil Steril 2022;118:101–8.
17. Fanton M, Nutting V, Rothman A, Maeder-York P, Hariton E, Barash O, et al. An interpretable machine learning model for individualized gonadotrophin starting dose selection during ovarian stimulation. Reprod Biomed Online 2022;45:1152–9.
18. Letterie G, Mac Donald A. Artificial intelligence in in vitro fertilization: a computer decision support system for day-to-day management of ovarian stimulation during in vitro fertilization. Fertil Steril 2020;114:1026–31.
19. Hariton E, Chi EA, Chi G, Morris JR, Braatz J, Rajpurkar P, et al. A machine learning algorithm can optimize the day of trigger to improve in vitro fertilization outcomes. Fertil Steril 2021;116:1227–35.
20. Chen T, Guestrin C. XGBoost. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM; 2016:785–94.
21. Groenewoud ER, Kollen BJ, MacKlon NS, Cohlen BJ. Spontaneous LH surges prior to HCG administration in unstimulated-cycle frozen-thawed embryo transfer do not influence pregnancy rates. Reprod Biomed Online 2012;24:191–6.
22. Ursillo L, Peyser A, Abittan B, Mullin C. A novel approach to natural frozen embryo transfers (FET). Fertil Steril 2021;116:e145.
23. Irani M, Robles A, Gunnala V, Reichman D, Rosenwaks Z. Optimal parameters for determining the LH surge in natural cycle frozen-thawed embryo transfers. J Ovarian Res 2017;10:70.
24. Bartels CB, Ditrio L, Grow DR, O'Sullivan DM, Benadiva CA, Engmann L, et al. The window is wide: flexible timing for vitrified-warmed embryo transfer in natural cycles. Reprod Biomed Online 2019;39:241–8.
25. Erden M, Polat M, Mumusoglu S, Ozbek IY, Dere GO, Sokmensuer LK, et al. Vitrified-warmed blastocyst transfer timing related to LH surge in true natural cycle and its impact on ongoing pregnancy rates. Reprod Biomed Online 2022;45:440–7.
26. Practice Committee of the American Society for Reproductive Medicine. Evidence-based treatments for couples with unexplained infertility: a guideline. Fertil Steril 2020;113:305–22.
27. Nulsen J, Wheeler C, Ausmanas M, Blasco L. Cervical mucus changes in relationship to urinary luteinizing hormone. Fertil Steril 1987;48:783–6.
28. Licht P, Fluhr H, Neuwinger J, Wallwiener D, Wildt L. Is human chorionic gonadotropin directly involved in the regulation of human implantation? Mol Cell Endocrinol 2007;269:85–92.

29. Fatemi HM, Kyrou D, Bourgain C, Van Den Abbeel E, Griesinger G, Devroey P. Cryopreserved-thawed human embryo transfer: spontaneous natural cycle is superior to human chorionic gonadotropin-induced natural cycle. Fertil Steril 2010;94:2054–8.
30. Litwicka K, Mencacci C, Arrivi C, Varricchio MT, Caragia A, Minasi MG, et al. HCG administration after endogenous LH rise negatively influences pregnancy rate in modified natural cycle for frozen-thawed euploid blastocyst transfer: a pilot study. J Assist Reprod Genet 2018;35:449–55.
31. Kyrou D, Kolibianakis EM, Fatemi HM, Grimbizis GF, Theodoridis TD, Camus M, et al. Spontaneous triggering of ovulation versus HCG administration in patients undergoing IUI: a prospective randomized study. Reprod Biomed Online 2012;25:278–83.
32. Kosmas IP, Tatsioni A, Fatemi HM, Kolibianakis EM, Tournaye H, Devroey P. Human chorionic gonadotropin administration vs. luteinizing monitoring for intrauterine insemination timing, after administration of clomiphene citrate: a meta-analysis. Fertil Steril 2007;87:607–12.
33. National Collaborating Centre for Women's and Children's Health. Principles of care. Available at: https://www.ncbi.nlm.nih.gov/books/NBK327792/. Accessed January 28, 2023.
34. Tiplady S, Jones G, Campbell M, Johnson S, Ledger W. Home ovulation tests and stress in women trying to conceive: a randomized controlled trial. Hum Reprod 2013;28:138–51.
35. Khattab AF, Mustafa FA, Taylor PJ. The use of urine LH detection kits to time intrauterine insemination with donor sperm. Hum Reprod 2005;20:2542–5.

La inteligencia artificial al servicio de la inseminación intrauterina y el coito programado en ciclos espontáneos

Objetivo: desarrollar un modelo de aprendizaje automatizado para predecir el momento de la ovulación y la ventana óptima para hacer una inseminación intrauterina o programar el coito (CP) en ciclos naturales.

Diseño: Estudio de cohorte retrospectivo.

Escenario: Unidad de fertilización in vitro grande.

Paciente(s): Pacientes que se llevaron a 2,467 ciclos de transferencia de embriones congelados en ciclo natural entre los años 2018 a 2022.

Intervención(es): Ninguna

Medida (s) de desenlace principal(es): Predicción de la exactitud del día óptimo para hacer la inseminación o el CP.

Resultado(s): El grupo de datos fue dividido en un grupo de entrenamiento que incluyó 1,864 ciclos, y dos grupos de prueba. En los grupos de prueba, la ovulación se determinó de acuerdo a la opinión de dos expertos en fertilidad independientes (“expertos”) (496 ciclos), o de acuerdo a la desaparición del folículo líder en ultrasonido dentro de dos días consecutivos (“certeza de ovulación”) (107 ciclos). Se usaron dos algoritmos: el modelo NGBoost de aprendizaje automatizado estimando la probabilidad de ovulación en cada día del ciclo y un algoritmo de manejo de tratamiento usando un modelo de aprendizaje para determinar el día óptimo de inseminación o si se debía usar un examen de sangre para determinarlo. Los niveles de estradiol, progesterona y hormona luteinizante fueron las características más importantes utilizadas por este modelo. El número promedio de exámenes fue 2,78 y 2,85 para el grupo de “certeza de ovulación” y de “expertos” respectivamente. En el grupo de “expertos”, el algoritmo predijo la ovulación correctamente y sugirió hacer la inseminación en el día 1 o 2 en el 92,9% de los casos. En el 2,9% el algoritmo predijo “pérdida”, significando que el día de la prueba, ya estaba sucediendo la ovulación o esta ya había pasado, sugiriendo no hacer la inseminación. En el 4,2% el algoritmo predijo un “error”, sugiriendo si hacer la inseminación, cuando en efecto se hubiese hecho en un día no óptimo (0 o -3). El grupo de “certeza de ovulación” tuvo resultados similares.

Conclusión (es): Hasta donde tenemos conocimiento, este es el primer estudio en implementar un modelo de aprendizaje automatizado, basado solamente en exámenes de sangre, para programar la inseminación o el CP con alta precisión, atribuido a la capacidad del algoritmo para integrar múltiples factores y no confiar solamente en el pico de hormona luteinizante. La introducción de las capacidades de este modelo, puede mejorar la exactitud y la eficiencia en la predicción de la ovulación e incrementar las posibilidades de concepción.

Número de Registro de estudio clínico: HMC-0008-21.