**Streamlining follicular monitoring during controlled ovarian stimulation? A data-driven approach to efficient IVF care in the new era of social distancing**

**I Robertson1\*, F P Chmiel2, Y Cheong3,4**

1. Human Development and Health, University of Southampton, Southampton UK, SO16 5YA, \*i.robertson@soton.ac.uk
2. IT Innovation Centre, School of Electronics and Computer Science, University of Southampton, UK, SO16 7NS
3. Human Development and Health, University of Southampton, Southampton, UK, SO16 5YA
4. Complete Fertility, Princess Anne Hospital, Coxford Road, Southampton

**Abstract**

**Study question:**

To evaluate the optimal follicular tracking strategy for controlled ovarian stimulation (COS) in order to minimise face-to-face interactions.

**Summary answer:**

Data from follicular tracking scans on day 8, 9 or 10 from the cycle start date can be used to make accurate predictions of trigger timing and risk of over response.

**What is known already**:

British Fertility Society guidance for centres restarting ART following Covid-19 pandemic related shutdowns recommends reducing the number of patient visits for monitoring during controlled ovarian stimulation (COS). Current evidence on optimal monitoring during ovarian stimulation is sparse, and protocols vary significantly. Small studies of simplifying IVF therapy by minimising monitoring have reported no adverse effects. There are opportunities to learn from the adaptations necessary during these extraordinary times to improve efficiency of IVF care in the longer term.

**Study design, size, duration:**

A retrospective database analysis of 9294 ultrasound scans performed during monitoring of 2322 IVF cycles undertaken by 1875 women in a single centre. The primary objective was to identify when in the IVF cycle data obtained from ultrasound is most predictive of both oocyte maturation trigger timing and an over-response to stimulation. If reduced frequency of clinic visits is needed, prioritising attendance for monitoring scans on the most predictive cycle days may be prudent.

**Participants/materials, setting, methods:**

Anonymised retrospective database analysis of IVF/ICSI cycles at a tertiary referral IVF centre in the United Kingdom. Machine learning models are used in combining demographic and follicular tracking data to predict cycle oocyte maturation trigger timing and over-response. The primary outcome was the day or days in cycle from which scan data yields optimal model prediction performance statistics. The model for predicting trigger day uses patient age, number of follicles at baseline scan and follicle count by size for the current scan. The model to predict over-response uses age and number of follicles of a given size.

**Main results and the role of chance**:

The earliest cycle day for which our model has high accuracy to predict both trigger day and risk of over-response is cycle day 8. The day 8 model to predict trigger date has MSE 1.70 +/- 0.11 and to predict over response AUROC 0.91 +/- 0.01.

**Limitations, reasons for caution**:

This is a retrospective single centre study and the results may not be generalisable to centres using different treatment protocols. The results are derived from modelling, and further clinical validation studies will further verify the accuracy of the model.

**Wider implications of the findings**:

Follicular tracking starting at day 8 may help streamline the amount of monitoring required in COS, a finding potentially useful in the new era of social distancing due to Covid-19. Previous small studies have shown that minimal monitoring protocols did not adversely impact outcomes. If IVF can safely be made less onerous on the clinic’s resources and patient’s time, without compromising success, this could help reduce burden related treatment drop-out.

**Study funding/competing interest(**s): F. P. Chmiel acknowledges funding from the NIHR *Applied Research Collaboration Wessex.*

**Introduction**

Traditionally, in vitro fertilisation (IVF) cycles involve patients attending repeat transvaginal ultrasound monitoring for follicular tracking during controlled ovarian stimulation. There are two crucial reasons for monitoring during COS, firstly to facilitate clinical decision making on the most appropriate timing to trigger final follicular maturation in order to maximize the number of mature oocytes and thus chance of live birth; and secondly, as part of a continued risk assessment of over response. Over-response puts patients at higher risk of OHSS, a rare iatrogenic complication of COS (Griesinger, et al., 2016). OHSS was the most common reported complication of ART in the 2015 European registry data with 2167 cases, an incidence rate of 0.44% (C De Geyter, 2020). In 2015-2016, 98 severe or critical OHSS cases were reported to the UK HFEA. Adequate monitoring is clearly required for patients triaged as high risk for this potentially life-threatening condition. Indeed, the OHSS working group consensus statement recommends ‘frequent vaginal ultrasonography and/ or serum oestradiol measurements’ for those identified to be at higher risk (1).

However, frequent monitoring visits are time consuming, costly and can be disruptive to patients’ daily routines, adding to both the practical and psychological burden of IVF. In a survey across four European countries, 21-36% of patients reported that difficulty fitting fertility treatments into their life and the need for repeated time off work are barriers to seeking treatment (Domar, et al., 2012). Frequent monitoring is also a time intensive use of clinic staff resources. The timing and frequency of follicular monitoring in COS is traditionally based on the clinical culture and individual preference and, in reality, the evidence for best practice is sparse. A 2014 systematic review of ultrasound for monitoring controlled ovarian stimulation concluded that more studies evaluating the optimal procedure for monitoring COS are needed, with a requirement for these to have live birth as the primary outcome and to be adequately powered (Kwan, et al., 2014). Aiming to reduce the burden of attending for repeated scans, some clinics have trialled home sonography as an alternative, but this technique has not been widely adopted (Gerris, et al., 2014, Gerris, et al., 2016). Simplification of IVF therapy by minimal monitoring has been reported to have no adverse effects on treatment outcome and the incidence of OHSS, although these studies were small and often used patient selection criteria and fixed dose protocols (Abdalla, et al., 1989, Hurst, et al., 2002, Roest, et al., 1995, Tan, 1994, Wikland, et al., 1994).

The objective of this study is to evaluate the optimal follicular tracking strategy for controlled ovarian stimulation (COS) in order to minimise face-to-face interactions in the era of social distancing. We use machine learning methods to predict the usefulness of each day of COS monitoring in assessing the timing of final follicular trigger and risk of over response. This evaluation aims to help streamline IVF scan schedules, whilst retaining the accurate predictive power needed for optimal, safe and personalised care.

**Materials and Methods**

Fully anonymised retrospective electronic data (IDEAS™, Mellowood Medical) on IVF and ICSI cycles was extracted from a tertiary IVF centre in the South of England, UK from 1/1/2011 to 30/11/2019. Most cycles are GNRH antagonist cycles (88.8%)*.* Fertility preservation, egg freezing, and altruistic egg donation cycles were excluded.

In this dataset, cycle day 1 is the first day of a patient’s start of menstruation. The usual practice for this centre includes a baseline transvaginal ultrasound scan on cycle day 2 to 4, with COS commencing subsequent to the baseline scan. Subsequent follicular tracking scan is then performed 5-7 days after the initial baseline scan, and then every other weekday until the maturation trigger is administered. Transvaginal oocyte collection is scheduled 36 hours following the trigger administration and performed under sedation.

Detailed patient data including demographics, cycle characteristics, antral follicle count (AFC), diagnosis, AMH (where available) and all follicle measurements taken during follicular tracking, medications and dosages used for stimulation, the type of oocyte maturation trigger and time/ date of trigger administration, outcome data (number of eggs collected, number of embryos frozen, live birth from fresh cycle, cumulative live birth from embryos created from this cycle) for each stimulation cycle were extracted from the database. Descriptive statistics were analysed for all variables in dataset.

Outcomes

The primary outcome was the machine-learnt model’s performance in predicting the day of trigger administration for predictions performed on each day of an IVF cycle and a separate model’s performance in predicting if the patient had over-responded and hence was at higher risk of OHSS. High response was defined as per the ESHRE ovarian stimulation guideline as more than 18 follicles ≥11 mm in size on day of oocyte maturation trigger and/or 18 oocytes collected (The ESHRE Guideline Group on Ovarian Stimulation, 2020). This definition was used as it has been shown to be associated with a significant risk increase in OHSS.(Griesinger, Verweij, Gates, Devroey, Gordon, Stegmann and Tarlatzis, 2016)

Machine Learning Methods

Random Forest Regressors were implemented as our predictive models using the sci-kit learn Python library (Pedregosa, et al., 2011). Models were trained using cross-validation at the treatment cycle level and the model’s performance was evaluated on the out-of-fold samples. Hyperparameter tuning was not performed to ensure we did not overfit the validation set, and reasonable model parameters were used (100 estimators in each model ensemble and no restriction on tree depth). Models were constructed for each cycle day using the results of individual ultrasound scans performed on the given day and combined with demographic data. Firstly, we constructed models to predict the day of trigger administration and, secondly, to identify the timing of scans most predictive of risk of OHSS. Missing values (less than 1 % of patients age was not recorded in the database) were imputed using mean imputation within respective cross-validation folds.

Statistical Analysis

The performance of the of the model in predicting the day of trigger administration was evaluated by calculating the mean squared error (MSE) between the predicted trigger administration day and the actual day of trigger administration. For comparison, this was compared to the accuracy of assuming all patients would be triggered at the mean trigger day for patients in the training set (also displayed as MSE vs patient trigger day).

An AUROC curve and model performance statistics for the model’s prediction of over-response is presented.

Ethical Approval

Ethical approval for this study was obtained from the University of Southampton ERGO II and NHS REC (IRAS Project ID: 275218). A data protection impact assessment was completed and approved by the University of Southampton DPIA panel on 23/01/2020.

**Results**

We collected data from follicular charts of 2322 cycles of 1875 women undergoing stimulation prior to oocyte retrieval for IVF, ICSI or oocyte donation. A total of 9294 individual scans are included in study. The mean age of patients included in the study was 33.58 years (range 20-44 years). 1505 patients had 1 cycle of COS within the dataset, 301 had 2 cycles and 55 had 3 cycles. After exclusion of oocyte donors and fertility preservation cycles, data from 2128 cycles from 1731 patients was complete and suitable for analysis.

Data presenting information on typical scan frequencies are presented in Figure 1(A&B). In Figure 1A, we present the proportion of treatment cycles that had a scan on a given day. Most patients have baseline scans on cycle days two or three, and scans are then performed on either cycle days four, five or six. Scans then occur with a high frequency between cycle days seven and twelve. Figure 1 B displays a histogram of the total number of scans each patient received in a single treatment cycle. The median number of scans for each patient was 4 (Interquartile range 3-5). A follicular growth chart for a patient with HCG trigger injection administered on cycle day 14 is displayed in Figure 1C, a typical visualisation used by clinical staff to help forecast the trigger administration day and a representation of the key data used by our predictive models.

Figure 2 displays a histogram of trigger administration day for each cycle in the dataset. In this dataset, the trigger injection was administered on mean day 11.9 +/- 1.41 day of the cycle.

In Figure 3, we display our machine-learnt models’ performance for predicting the day of trigger administration for each scan day. Models using data from any baseline scan are not highly predictive of the trigger administration day (the mean MSE of models using data only from baseline scans is 3.83 +/- 0.24) and do not perform significantly better than simply assuming each patient triggers on the mean day of trigger administration of historical cycles (dashed line in Figure 3). However, models built using data from scans performed later in the cycle become much more predictive and significantly outperform the baseline trigger day prediction (compare dashed and solid lines in Figure 3 for days six and beyond).

This is further supported by looking at the model’s predictions in detail, as shown in Figure 4 for several representative days. For follicular data obtained from baseline scans (with little discernible follicle growth), the model cannot reliable stratify patients into their eventual trigger administration day and does not perform better than simply assuming each person will trigger on the historical mean of trigger administration day (Figure 4A). By day 6 (Figure 4B) the model can predict with reasonable confidence patients triggering time, such that is can reliably assign patients into groups segregated by the expected imminence of trigger administration. At day 8 the model is strongly predictive of the day of trigger administration (data points follow the grey line representing a perfect predictor in Figure 4C much more closely). The predictive ability of this model suggests that cycle days 8-10 are most useful in predicting trigger timing and earlier scans hold significantly less benefit in forecasting the trigger administration day.

The performance of our model trained to predict patients who over-respond to treatment is displayed in Figure 5, where we plot the AUROC (evaluated on the out-of-fold samples) for each model trained using scan data collected from the respective scan day. For baseline scans (cycle days 2 to 4), the model can predict over-response with moderate accuracy (mean AUROC 0.77 +/- 0.01), showing reasonable predictions of over-response can be made using the number of follicles present at the baseline scan and patient age. As the treatment cycle progresses over-response can be predicted with increasing accuracy as the follicles grow and the model can begin to predict over response with high precision and recall (AUROC exceeding 0.90) beyond cycle day 8.

**Discussion**

The principle findings of this study are that follicular tracking data obtained from cycle days 8-10 can accurately predict both the timing of ovulation trigger and the chance of over response. Our results suggest that it may be possible to reduce the number of COS monitoring visits, without compromising predictive power to identify timing of egg collection or increasing the risk of over response, by starting follicular tracking from cycle day 8. Our model suggests that follicular tracking early in stimulation offer little insight into cycle progress. However, seven patients in this dataset (0.33%) had the trigger administered prior to cycle day 8 and would experience delayed trigger with this proposed tracking approach. The mean leading follicle size at the last scan pre-trigger for these 7 patients was 18mm and it is possible that a delay of one day is of little clinical significance to their cycle outcome.

Generally, the patient’s age and ovarian reserve (AMH and /or antral follicle count) are used to triage their risk of over and under-response to COS. This study demonstrates that additional predictive power is gained from follicular tracking data, particularly when the scan is performed on day 8-10, compared to earlier scans. Accurate prediction of over-response on day 8 is useful to assist with decision making regarding use of a GnRH agonist trigger, with time for prescription and ordering as necessary, and time for counselling, including regarding use of a freeze all embryo approach.

Many clinicians perform a baseline scan early in the cycle to assess antral follicle count, check endometrial thickness (ET) and assess for presence of ovarian cysts or endometrial polyps that could warrant postponement or cycle cancellation. There is a lack of evidence for benefit of the latter, particularly as most patients have had a detailed, often including 3D ultrasound scan as part of their pre-IVF work-up. Delaying stimulation start for a thick ET at baseline scan is not justified by robust evidence and ESHRE guidelines recommend measuring one ET at the time of trigger scan only (The ESHRE Guideline Group on Ovarian Stimulation, 2020). Sensibly streamlining COS follicular tracking protocols would reduce the number of patient contacts per cycle, potentially increasing safety for patients and clinic staff.

Strengths and Weaknesses

Not every patient has a scan on each day, this has the effect of the models test set for each scan day being different. The impact of this is mitigated by using cross-validation, which we use to evaluate the model performance on unseen data and provide an estimate of the confidence intervals of the performance metrics. Ideally, a prospective study would be performed where a daily scan is performed to further validate this work.

This is a retrospective single centre study and our models would benefit from further validation by testing their generalization to external data (i.e., collected at different centres). To this end, our models could be readily applied to fully anonymised data easily extractable (following ethical approval) from the IDEAs database of other centres to test the present findings.

We emphasise that non-inferiority of reduced follicular tracking frequency cannot be surmised from these retrospective findings. As in all clinical research, it is best practice that prior to any change in routine practice, randomised trial of any streamlined protocols should be carried out in comparison to routine care, with core outcome sets reported, a primary outcome of healthy live birth and key secondary outcomes including the number of cases of OHSS and validated measures of patient’s psychological wellbeing.

The trigger day targeted by these models is the timing used for each individual patient, as decided by daily clinical multidisciplinary review of follicular growth charts. This may not have been the optimal timing to maximise cycle outcome. Confounders will include physician tendencies/bias, avoidance of weekend scans and oocyte pick-up and the fact that each patient’s treatment was individually tailored. In addition, this study was performed within a single centre operating within its standard operating protocols. Results may not be generalisable to other centres that, for example, significantly base decision making on blood monitoring during stimulation or have a higher rate of within-cycle stimulation dosage changes. AMH result was not available for a significant proportion of patients in this database and was not included in our modelling.

**Conclusion**

The finding of good predictive power for both trigger timing and over-response risk from single scans on day 8, 9 or 10 agrees with previous small studies suggesting that reducing follicular tracking scan frequency may be justifiable. In the wake of the current COVID-19 pandemic, if there is a significant need to reduce patient contact per cycle in order to facilitate safe access to care, priority for patients attending follicular tracking scans after day 7 of the treatment cycle should be considered. In future, it is possible that combining validated predictive models with follicle tracking ultrasound data may make further streamlining of monitoring possible.

Our current models require further testing in external validation studies. Nonetheless, we anticipate with ongoing development, the model will prove a useful aid both for clinics scheduling their upcoming procedures and for clinicians making crucial ovulation trigger timing decisions. We anticipate that testing to compare model predictions with expert physician estimates of trigger timing will be a key step to evaluate model performance prior to clinical implementation.

**Author’s Roles**

Dr Isla Robertson envisaged study, performed data curation and analysis, contributed to machine-learning modelling and led manuscript writing. Dr Francis P Chmiel performed data analysis, led machine-learning model development and contributed to manuscript writing. Professor Ying Cheong provided clinical insight, contributed to manuscript writing and supervised project. All authors approved the final version to be published.

**Data Availability Statement**

The data and code underlying this article will be shared on reasonable request to the corresponding author.

**References**

Abdalla HI, Baber RJ, Leonard T, Kirkland A, Mitchell A, Power M, Owen E, Studd JWW. Timed oocyte collection in an assisted concepton program using GnRH analog. *Human Reproduction* 1989;4: 927-930.

C De Geyter CC-J, M S Kupka, C Wyns, E Mocanu, T Motrenko, G Scaravelli, J Smeenk, S Vidakovic, V Goossens ART in Europe, 2015: results generated from European registries by ESHRE. *Humn Reproduction Open* 2020;2020.

Domar A, Gordon K, Garcia-Velasco J, La Marca A, Barriere P, Beligotti F. Understanding the perceptions of and emotional barriers to infertility treatment: a survey in four European countries. *Human Reproduction* 2012;27: 1073-1079.

Gerris J, Delvigne A, Dhont N, Vandekerckhove F, Madoc B, Buyle M, Neyskens J, Deschepper E, De Bacquer D, Pil L *et al.* Self-operated endovaginal telemonitoring versus traditional monitoring of ovarian stimulation in assisted reproduction: an RCT. *Human Reproduction* 2014;29: 1941-1948.

Gerris J, Vandekerckhove F, De Sutter P. Outcome of one hundred consecutive ICSI attempts using patient operated home sonography for monitoring follicular growth. *Facts Views and Vision in Obgyn* 2016;8: 141-146.

Griesinger G, Verweij PJM, Gates D, Devroey P, Gordon K, Stegmann BJ, Tarlatzis BC. Prediction of Ovarian Hyperstimulation Syndrome in Patients Treated with Corifollitropin alfa or rFSH in a GnRH Antagonist Protocol. *Plos One* 2016;11.

Hurst BS, Tucker KE, Schlaff WD. A minimally monitored assisted reproduction stimulation protocol reduces cost without compromising success. *Fertility and Sterility* 2002;77: 98-100.

Kwan I, Bhattacharya S, McNeil A. Monitoring of stimulated cycles in assisted reproduction (IVF and ICSI). *Cochrane Database of Systematic Reviews* 2014: N.PAG-N.PAG.

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V *et al.* Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 2011;12: 2825-2830.

Roest J, Verhoeff A, Vanheusden AM, Zeilmaker GH. Minimal monitoring of ovarian hyperstimulation- A useful simplification of the clinical-phase of an in-vitro fertilzation treatment. *Fertility and Sterility* 1995;64: 552-556.

Tan SL. Simplifying In-vitro fertilizaton therapy. *Current Opinion in Obstetrics & Gynecology* 1994;6: 111-114.

The ESHRE Guideline Group on Ovarian Stimulation EB, Simone Broer, Georg Griesinger, Michael Grynberg, Peter Humaidan, Estratios Kolibianakis, Michal Kunicki, Antonio La Marca, George Lainas ESHRE guideline: ovarian stimulation for IVF/ICSI. *Human Reproduction Open* 2020;2020.

Wikland M, Borg J, Hamberger L, Svalander P. Simplificaton of IVF - Minimal monitoring and the use of subcutaneous highly purified FSH administration for Ovulation Induction. *Human Reproduction* 1994;9: 1430-1436.