O-185   A clinically robust machine learning model for selecting the first FSH dose during controlled ovarian hyperstimulation: incorporating clinical knowledge to the learning process

N. Correa Mañas[1], J. Cerquides[2], J.L. Arcos[2], R. Vassena[1], M. Popovic[1]

[1]EUGIN group, Research and Development, Barcelona, Spain
[2]IIIA-CSIC, Learning Systems, Bellaterra, Spain

Study question: Can we improve the selection of the first dose of follicle stimulating hormone (FSH) for IVF patients by incorporating clinical evidence with machine learning?

Summary answer: A significant improvement in correct dose prescription is achieved when using a FSH dosing model that integrates a novel machine learning methodology with clinical knowledge.

What is known already: The appropriate selection of the initial FSH dose during controlled ovarian hyperstimulation (COH) is key for ART success. An optimal number of oocytes should be obtained for achieving pregnancy, while also avoiding complications for the patient. While clinical protocols achieve good results for the majority of patients, further refinements in individualized FSH dosing may improve safety while reducing the risk of poor response. Machine learning techniques have been applied to dose/response problems with promising results. However, the observational datasets used to train the models are often incomplete and biased, leading to less clinically robust predictions.

Study design, size, duration: We used a database of 2713 first cycles performed between January 2011 and December 2019, across five clinics to train the model. Predictors included age, body mass index (BMI), anti-Müllerian hormone (AMH) levels and antral follicle count (AFC). First FSH dose and number of mature oocytes retrieved were also recorded. Another 273 cycles performed between January 2020 and September 2021 were used for validation. Dosing was performed by 41 clinicians (mean 12 years of experience).

Participants/materials, setting, methods: The model was developed in Python and trained while incorporating available clinical evidence. The relationship between FSH dose and response (number of mature oocytes) was specified by assuming a positive linear function. Rules were incorporated to penalize dose changes that resulted in poorer outcomes compared to the clinician, and to discourage large dose modifications by the model. These same rules were applied during the validation phase. Comparisons were assessed by Chi-squared test with Bonferroni correction.

Main results and the role of chance: Mean maternal age was 37.1 ± 4.9 years. Patients had a mean BMI of 23.8 ± 4.2, AFC of 11.9 ± 7.7 and AMH of 2.4 ± 2.3. They were prescribed a first dose of $247.0 \pm 59.0\,IU$ of FSH and obtained 7.3 ± 5.3 mature oocytes after pick-up. An optimal outcome was defined as 10-15 mature oocytes retrieved, and 300 IU was the maximum FSH dose allowed. In the validation dataset, we observed that 22.9% patients could have received a higher FSH dose (<10 mature oocytes collected), 69.3% were correctly prescribed (10-15 mature oocytes or no change in dose was required), while 7.9% needed a decrease in FSH dose (>15 mature oocytes). In the machine learning model 9.4% of patients required a dose increase, 80.5% were correctly prescribed and 7.9% needed a decrease in dose. Accordingly, the model rescued 11.6% of patients that would have otherwise achieved suboptimal outcomes. Ultimately, when compared to the clinicians, the model significantly reduced the number of patients that required a dose increase, and significantly improved the number of patients that were prescribed a correct dose ($p < 0.05$).

Limitations, reasons for caution: The model assigned a non-conservative dose increase to 2.3% of patients, potentially leading to a risk of ovarian hyperstimulation syndrome. Moreover, hyper-responders were underrepresented in the database, resulting in no significant improvements in the dose/outcome relationship for this subgroup.

Wider implications of the findings: Coding clinical knowledge into the training of clinically relevant machine learning models ensured adherence to scientific evidence, and improved model performance while making conservative changes. This approach is critical for the safe and robust clinical implementation of interventional models.

Trial registration number: not applicable