# Scientific and Clinical Advances Advisory Committee (SCAAC) – Matters arising

## Monday 3 October 2022

| Date | Action | Responsibility | Due date | Progress to date |
|---|---|---|---|---|
| 31/01/2022 | Assess whether further outputs are required in the topic of the impact of the microbiome, and whether it needs to be considered as a treatment add-on. | Victoria Askew, Policy Manager | Ongoing | This will be assessed as part of an agenda item at the June 2023 SCAAC meeting. This has been amended from the SCAAC workplan due to internal resourcing restrictions. |
| 06/06/2022 | The Committee will continue to monitor and share relevant literature regarding public health impacts on fertility, assisted conception and early pregnancy more generally. | All SCAAC members | Complete | A standing agenda item 'Relevant public health developments' is due to be discussed at this meeting as agenda item four. |
| 06/06/2022 | The impacts of stress on fertility treatment outcomes, and more specifically potential stress management tools, should remain as a medium priority topic of the SCAAC and be brought back to the committee for consideration at a future meeting. | All SCAAC members | Complete | The Executive will consider this recommendation when creating the SCAAC workplan for 2023/24, to be presented to the Committee at the February 2023 meeting. |
| 06/06/2022 | Following discussions and decisions regarding the application of the addition of Androgen supplementation as a treatment add-on. Members expressed concern over language used | Sonia Macleod, Scientific Policy Manager | Ongoing | The Executive will amend the treatment add-ons application form decision tree in line with the evolving treatment add-ons rating system. This will be |

| | | | | |
|---|---|---|---|---|
| | within the treatment add-ons eligibility criteria. With the Authority considering possible changes to both the evidence base and how evidence is presented, members requested for the decision to be reviewed and presented to the Committee at a future meeting. | | | presented at the February SCAAC meeting. |
| 06/06/2022 | The Committee made a recommendation to the Authority that in the absence of good and robust RCTs or meta-analyses, expanding the evidence base may be necessary and helpful when assigning treatment add-on ratings. | Sonia Macleod, Scientific Policy Manager | Complete | This recommendation was taken to the Authority meeting in July. A decision tree on the evidence base will be presented to the Committee today under the Add-ons agenda item. |

# Treatment add-ons rating system review – an update

## Details about this paper

| | |
|---|---|
| Area(s) of strategy this paper relates to: | The best care/The right information |
| Meeting: | Scientific and Clinical Advances Advisory Committee (SCAAC) |
| Agenda item: | 6 |
| Paper number: | HFEA (03/10/2022) 006 |
| Meeting date: | 03 October 2022 |
| Author: | Sonia Macleod, Scientific Policy Manager |
| Annexes | Annex A: The new add-ons rating system<br>Annex B: Draft Evidence Decision tree for SCAAC rating of add-ons |

## Output from this paper

| | |
|---|---|
| For information or recommendation? | The SCAAC is asked:<br>• for member's views on the evidence decision tree for rating add-ons<br>• whether ongoing pregnancy could be considered a proxy for live birth rate |
| Recommendation: | NA |
| Resource implications: | NA |
| Implementation date: | NA |
| Communication(s): | User Acceptance Testing on the redesigned add-ons web pages will start shortly and a full communications plan to engage patients and clinics will accompany the launch of the updated webpages in Spring 2023. |
| Organisational risk: | Medium |

## Introduction

**1.1.** Addressing how treatment add-ons are offered by clinics and information given to patients is a key feature of the HFEA strategy for 2020-24.

**1.2.** Our work on evolving the traffic light system for rating some treatment add-ons has had three key elements

- Evolving the rating scale and presentation of this evidence
- Incorporating outcomes in addition to live birth rates
- Determining the evidence base which should be used by SCAAC when rating add-ons.

**1.3.** At the Authority meeting in July 2022 it was agreed that the add-ons rating system would evolve as set out in Annex A. Briefly this included

- Moving to a five category rating scale, set out below
- Including outcomes in addition to live births.
- Expanding the evidence base in line with the recommendation made by SCAAC in June 2022
- Consequential changes to the definition of an add-on which HFEA use when determining whether to rate an add-on.

**1.4.** This paper outlines these changes and how they will impact on the way SCAAC rates add-ons

## SOP and Decision tree to determine add-on eligibility for HFEA rating

**1.5.** An SOP will be developed for SCAAC to use when rating add-ons. This will contain two decision trees. The first decision tree will be used as part of **the application process** to determine whether an add-on is one that SCAAC should give a rating to. The second decision tree will define the evidence to be used when rating eligible add-ons.

**1.6.** We are currently developing an decision tree to determine if an add-on is one that SCAAC should give a rating to based on the decisions made at the July Authority meeting where consequential changes to the definition of an add-on were agreed.

**1.7.** If in the future other regulatory bodies also provide patient information ratings on add-ons then we will consider options, which could include continuing to provide such information ourselves, collaborating with others to provide a unified source of information, or signposting to the information that others provide.

## SOP and evidence decision tree for eligible add-ons

**1.8.** Both the SCAAC recommendation and the Authority decision on the evidence base emphasised that any HFEA process should align with the processes used by similar organisations such as NICE, Cochrane and MHRA. The draft evidence decision tree has been designed with this in mind. This can be found in Annex B,

**1.9.** Live birth rates will remain the primary focus of the HFEA add-ons rating system. Members are asked whether they consider ongoing pregnancy could be used as a proxy for LBR where there is no LBR data available. This happens at other organisations, for example the European Medicines Authority accept this for the follitropin delta registration trial due to the close correlation with live birth. This would potentially widen the evidence base. Consideration would be needed of how to carefully communicate this difference to patients.

**1.10.** The SOP requires an initial decision on identifying which populations and/or outcomes other than LBR are or interest for each add-on. It is proposed that this is carried out by:-

- the Chair of SCAAC;
- one person from the HFEA who is either a member of the scientific policy team or is a member of the Register Research Panel;
- at least one clinician; and
- at least one person who is involved in clinical research/embryology.

Once these outcomes/populations have been identified the scientific policy team at the HFEA will conduct literature searches to identify relevant publications.

**1.11.** Identified publications will be sent for external review by an appropriate expert using GRADE methodology as part of a quality control step.

**1.12.** The draft decision tree mirrors NICE's requirement for a minimum of three publications.

**1.13.** Although the Authority proposed that evidence should be 'high quality' defined using GRADE methodology. This has been amended in the draft decision tree to at least 'medium quality' evidence on the advice of SCAAC members due to the scarcity of 'high quality' evidence.

**1.14.** The decision tree sets out the prioritisation of different evidence types, and reflects the recommendation from the June 2022 SCAAC meeting that RCTs should remain the preferred evidence base, but where high quality RCTs are not available non-RCT evidence should be considered.

**1.15.** Where there are not sufficient publications of the required quality an add-on will be rated grey. A grey rating does not preclude the HFEA from publishing any other information that is known on each add-on. For example, if there was just one very strong RCT indicating patient benefit this could be included on the section on that particular add-on on the HFEA website.

**1.16.** If there are sufficient medium or high quality publications then SCAAC will rate the add-ons, informed by the report of the external reviewer.

## Next Steps

**1.17.** Up to this point SCAAC have been reviewing add-ons on an annual basis at their October meeting. Expanding the evidence base may mean this approach is no longer sustainable and could lead to longer intervals between ratings, and/or to reviewing the rating for different add-ons at different times of the year rather than the current system where all add-ons are reviewed together once a year. This will be reviewed by HFEA once the preparation for reviewing add-ons ratings at the February SCAAC meeting has been completed.

**1.18.** **The draft SOP for rating add-ons, including the evidence decision tree, will be agreed and adopted.** Add-ons meeting the requirements set out in the SOP will be reviewed at the February 2023 SCAAC meeting.

**1.19.** **The presentational aspects of the new add-ons system will be subject to user acceptance testing.** In parallel with the work on the SOP, user-acceptance testing of the proposed changes to the HFEA website will be undertaken.

**1.20.** In Spring 2023 the updated ratings will be launched on the HFEA website **with an accompanying communications plan to ensure clinic staff and patients are aware of the new ratings system**.

## Recommendations

**1.21.** SCAAC is asked:

- for member's views on the evidence decision tree for rating add-ons,
- whether they consider ongoing pregnancy could be considered a proxy for LBR

# Annex A – Authority Decisions on evolving the add-ons rating system

- Below are the options approved by Authority in July 2022.

## Five Category rating system

- This is a change to the current three category RAG (red, amber, green) rating system.
- **The Authority approved the following options and the wording.**

On balance, the evidence from high quality studies shows **this add-on is effective** at improving treatment outcomes for most fertility patients.

On balance, **it is not clear whether this add-on is effective** at improving treatment outcomes for most fertility patients. This is **because there are conflicting findings** between different high quality studies – in some studies the add-on has been found to be effective, but in other studies it has not.

**We cannot rate the effectiveness of this add-on** at improving treatment outcomes for most fertility patients as there have been so few or no studies done.

On balance, the evidence from high quality studies shows that **this add-on has no effect on treatment outcomes** for most fertility patients.

**There are potential safety concerns** and/or, on balance, the evidence from high quality studies show that **this add-on may reduce treatment effectiveness for most fertility patients.**

## Additional outcomes

- This would mean that the above five category system will be used to rate outcomes such as miscarriage and outcomes for specific patient groups, for example those over 35, in addition to live births.
- Authority have agreed additional outcomes in addiction to live births and have tasked SCAAC with determining which outcomes should be rated for each add-on.

## Expanding the Evidence Base

- Authority have agreed to expand the evidence based used to rate add-ons in line with SCAAC's recommendation that in the absence of high-quality RCTs or meta-analysis the evidence base should be expanded.

- Any change to the evidence base should be broadly aligned to the methodology already used by Cochrane, NICE and MHRA.
- Authority agreed that a decision tree/algorithm should be developed with input from SCAAC and an expert statistician. This is underway and will be used in conjunction with

## Consequential changes

- HFEA currently provide information on add-ons that meet the following criteria.
    - Additional treatments (to the core treatment e.g. IVF or IUI), that patients need unbiased information about effectiveness and risks, that are being offered in fertility clinics;
    - where there is published scientific literature of a good RCT investigating the treatment's ability to improve the chances of having a baby; and
    - where evidence on efficacy or safety for the use of the treatment in a clinical setting is lacking or absent.

- The following amendments were proposed.
    - Additional treatments (to the core treatment e.g. IVF or IUI) that are being offered to the general patient population in licensed fertility clinics in the UK,
    - Where there is published scientific literature which claims to demonstrate that the add-on improves live birth rates or other treatment outcomes rated by the HFEA; but
    - where evidence of effectiveness for the use of the treatment in a clinical setting is lacking or absent; and
    - where patients need unbiased information about the effectiveness and risks of this treatment.

- The Authority agreed consequential changes to these criteria, subject to input from the Chair of SCAAC on points three and four

# Annex B – Draft Evidence decision tree for SCAAC to use when rating add-ons

## Evidence priority rankings (taken from NICE):

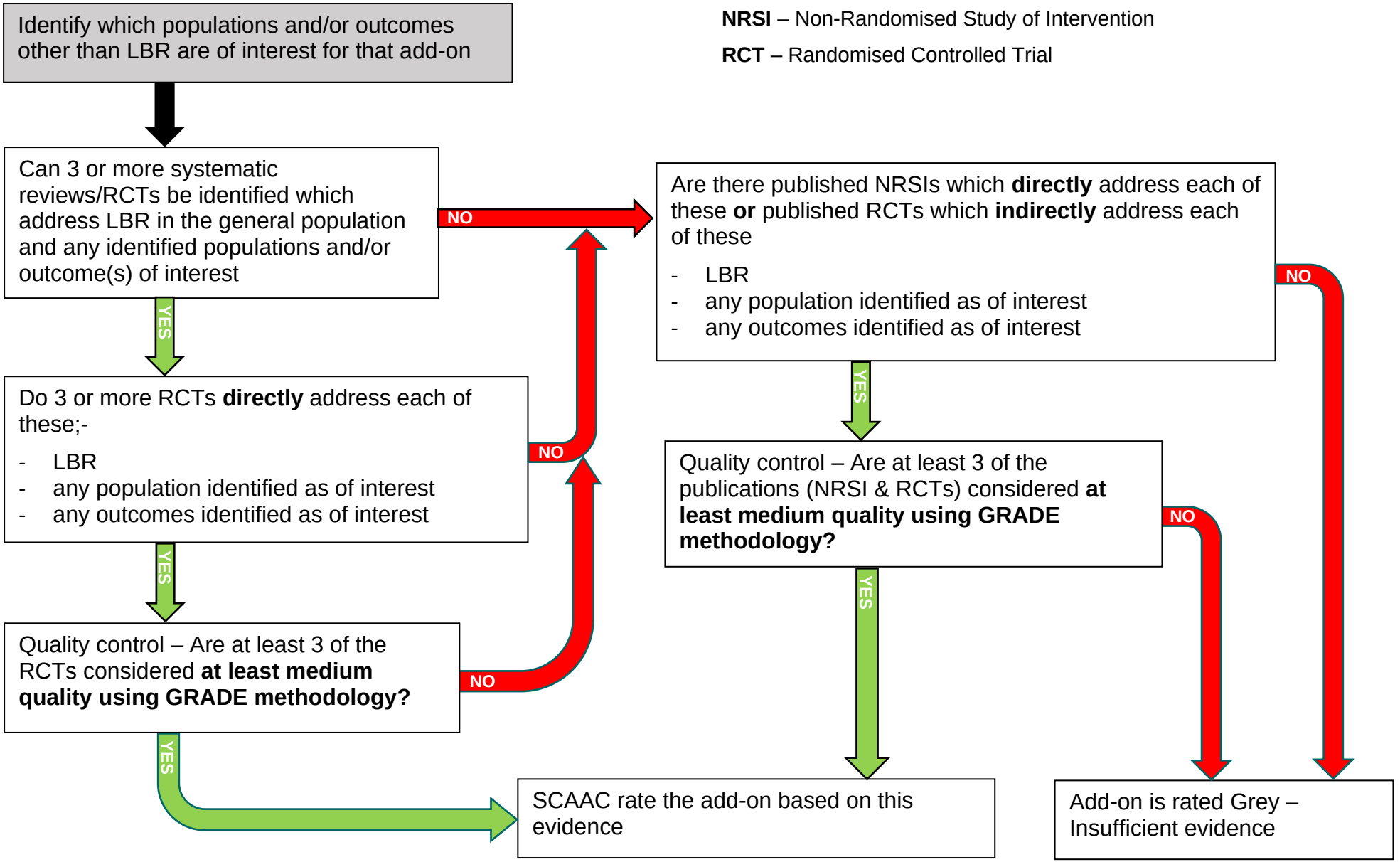
NICE chose the **top three pieces of evidence prioritising**:-

1. Systematic reviews
2. Randomised control trials

Cohort/case-control/case series, ranked upon a combination of their size/publication date/clarity of data/inclusion of an “active comparator” (effectively, a placebo option)/how representative the study population is of the relevant

When applying this to add-ons if none of the above can be identified, the intervention will be rated grey – **We cannot rate the effectiveness of this add-on** at improving treatment outcomes for most fertility patients as there have been so few or no studies done

## Evidence decision tree for eligible add-ons



# Artificial Intelligence.

## Details about this paper

| | |
|---|---|
| Area(s) of strategy this paper relates to: | Shaping the future |
| Meeting: | Scientific and Clinical Advances Advisory Committee (SCAAC) |
| Agenda item: | 8 |
| Paper number: | HFEA (03/10/2022) 008 |
| Meeting date: | 03 October 2022 |
| Author: | Victoria Askew, Policy Manager |

## Output from this paper

| | |
|---|---|
| For information or recommendation? | For recommendation |
| Recommendation: | Members are asked to:<br>• advise the Executive if they are aware of any other recent developments in AI relevant to fertility treatments or research.<br>• discuss their views of the impact AI will have on fertility treatment as technology advances, including practical and ethical challenges to their application.<br>• discuss their views on the Authority's regulatory interest around AI systems, scope, and limits, particularly considering the Office for AI's regulation of AI policy position paper.<br>• review whether any outputs from the HFEA are required, addressing the use, or regulation, of AI. |
| Resource implications: | Dependant on evaluation of HFEA remit in relation to AI regulation. |
| Implementation date: | NA |
| Communication(s): | Short meeting summary to be published in Clinic Focus. |
| Organisational risk: | Medium |

# 1. Background

**1.1.** Artificial intelligence (AI) is the theory and development of computer systems able to perform tasks normally requiring human intelligence, typically making predictions or decisions such as visual perception, speech recognition, decision-making, and translation between languages.

**1.2.** AI is driven by data. In healthcare, this could be related to data pertaining to patient characteristics or data from medical images. With a large enough dataset, machine learning can be applied to create algorithms independently and form systems such as artificial neural networks (ANNs) that are advanced enough to generate clinical judgments or predictions.

**1.3.** Within reproductive medicine, the potential application of AI continues to expand. Uses include supporting clinical decision-making, predicting patient outcomes and success rates, grading or selection of sperm, eggs and embryos, analysis of pre-implantation genetic testing for aneuploidy (PGT-A) samples, as well as the non-invasive PGT-A, through to the analysis of videos, static images and spent culture media.

**1.4.** There are issues that need to be considered with the introduction of AI-driven processes into clinical practice. In summary, it is not always possible to explain how decisions are made by machine learning models. This lack of transparent decision-making creates both legal and ethical concerns and could risk creating unintentionally biased decisions. Training AI systems requires large amounts of data to create high-quality and reliable outputs. Considerations also need to be made for obtaining informed consent for sharing personal data and considering the implications of data passing between countries and the accountability of each element of a model's output. Concerns also arise from how to manage the resource implications for a potential increase in demand for HFEA data to develop AI systems.

**1.5.** The regulation of AI is a topic that the HFEA has been concerned with for some time. The committee last discussed AI at the June 2021 SCAAC meeting. Members discussed the need to focus on what the technology is being used for, particularly what claims clinics are making to patients and how it is being sold and charged for. It was noted that it might be necessary to consider definitions and to limit the scope of this work.

**1.6.** The committee summarised that the rapid development of AI is a challenge for all healthcare regulators. At the HFEA, the inspectors, in particular, may find it difficult to analyse algorithms and their results. This is because AI-driven technology is already being sold to patients as an advanced tool, greater than relying on a clinician's experience, but not all data behind algorithms are being published. Commercial and academic organisations are approaching the HFEA for access to register data; some of them have no intention of publishing how their algorithm is developed. The underlying data determines how effective embryo grading models are for each clinic. There may be flaws in AI models that use historical data, if there is variation in policies and practices over that time.

# 2. Developments

### Informing fertility treatment pathways

**2.1.** AI could be used as a tool to help clinicians make recommendations about different aspects of a patient's fertility treatment pathway. Different AI models have been created that use patient characteristics to predict their chances of having a live birth or clinical pregnancy following

fertility treatment, with overall moderate performances to date (Bardet et al., 2022; C. N. Barreto et al., 2022; Khodabandelu et al., 2022; Wang et al., 2022; Zhang et al., 2022). These prediction tools could allow patients to receive accurate, personalised success rates and agree with their clinician on how and when to proceed with fertility treatment. They could also be used as a training tool or as part of the clinic's quality management system.

**2.2.** Of note, fertility prediction tools are already publicly available. This includes the **Society for Assisted Reproductive Technology (SART)**, **Apricity**, **Univfy** or the **Centre for Disease Control and Prevention (CDC)**. When given the results, patients may be encouraged to make lifestyle changes to increase their chances of success, or the results can inform payment plans for fertility treatment, such as refund or multi-cycle programmes.

**2.3.** Multiple factors may play a role in a patient's chances of a successful treatment outcome. Beyond overall success rates, AI could also be used to create more specific tools for clinicians to make decisions about certain aspects of a patient's treatment.

**2.4.** A study by Mehrjerd et al., 2022 looked at 729 couples with unexplained infertility to analyse the impact of endometrial thickness on ongoing pregnancy rate. The group used a random forest model and logistic regression to predict pregnancy following intracytoplasmic sperm injection (ICSI), in vitro fertilisation (IVF) and intrauterine insemination (IUI) treatments. The study found an endometrial thickness cut-off for ongoing pregnancy of 7.7mm for IUI and 9.99mm for IVF or ICSI.

**2.5.** Wen et al., 2022 used data from 1507 fresh embryo transfer cycles to build six machine learning algorithms to predict pregnancy outcomes and multiple pregnancy risk associated with the number of embryos transferred. The pregnancy prediction model produced accuracy of 0.716, sensitivity of 0.711, specificity of 0.719, and area under the curve (AUC) of 0.787. The multiple pregnancy prediction model produced an accuracy of 0.711, sensitivity of 0.649, specificity of 0.740, and AUC of 0.732. The authors concluded that the AI models provide reliable outcome prediction and could be a promising method to decrease multiple pregnancy risk after IVF.

**2.6.** A study by Shen et al., 2022 aimed to determine the optimal number of embryos to transfer in patients who had experienced recurrent implantation failure whilst reducing the risks of experiencing a multiple pregnancy. The group used HFEA data to develop four machine learning algorithms with two groups of patients. Group A included 34,175 cycles of treatment with two embryos transferred. Group B included 11,746 cycles of treatment with one embryo transferred. The AdaBoost model of Group A obtained the best performance, while the GBDT model in Group B was proved to be the best model. Both models showed the potential to provide accurate predictions of transfer outcomes.

**2.7.** Correa et al., 2022 undertook an observational study of patients from five fertility clinics to identify the optimal dose of FSH in ovarian stimulation. 2,713 patients were used to develop the algorithm, and it was tested on 774 patients. The model reached a mean performance score of 0.87 in the development phase, significantly better than for doses prescribed by clinicians for the same patients (0.83). The mean performance score of the model recommendations was 0.89 in the validation phase, also significantly better than clinicians (0.84). The authors concluded that the model was shown to surpass the performance of standard practice.

**2.8.** It is important to consider the diversity of patients within a data set. If the data used to train and validate an AI model is limited in terms of size or geographical distribution, it could affect the generalisability of the prediction tool created (Abdullah et al., 2022). There are already well-

publicised concerns around the potential for discrimination and bias in AI. This includes those outlined in the **Centre for Data Ethics and Innovation's** (CDEI) review into bias in algorithmic decision making and Imperial College London's report **'Addressing racial and ethnic inequalities in data-driven health technologies'**. It is important to ensure that any prediction tools would not exacerbate existing health inequalities, such as those outlined in the HFEA **ethnic diversity report**, which looked into how access to and outcomes of fertility treatment differ by ethnic group.

### Increased efficiency

**2.9.** Letterie et al., 2022 analysed the data of 1,591 patients to design an algorithm that could minimise clinic visits and improve workflow during ovarian stimulation. The algorithm was able to identify the best day for monitoring, with a mean error of 1.355 days. After determining a monitoring day, a trigger date and range of three oocyte retrieval days were specified. Accuracy for predicting the total number of oocytes with baseline testing alone or in combination with data on the day of observation was 0.76 and 0.80, respectively. The sensitivities for estimating the total number and number of mature oocytes based solely on pre-IVF profiles in group one (0-10) were 0.76 and 0.78, and in group two (>10) 0.76 and 0.81, respectively.

**2.10.** Hammer et al., 2022 used data from 4,889 time-lapse embryo images to develop an AI witnessing algorithm. The convolutional neural network (CNN) processed embryo images for each patient and produced a unique identification key associated with the patient ID on day 3 and day 5, forming a data library. The algorithm then evaluated the embryos at a later time point on days 3 and 5 and generated another key that was matched with the patient's unique key in the library. This was then tested using 400 patient embryo cohorts on days 3 and 5. The CNN matched the patient identification within random pools of 8 patient embryo cohorts on day 3 with 100% accuracy (n = 400 patients; 3 replicates). For day 5 embryo cohorts, the accuracy within random pools of 8 patients was 100% (n = 400 patients; 3 replicates).

**2.11.** A review by Abdullah et al., 2022 discussed developments in automation, with and without AI, within the fertility sector. This included the use of AI trained using ultrasound images to detect empty follicles and oocyte containing follicles at egg collection, which could increase accuracy and decrease duration of transvaginal oocyte retrieval. The review also discussed the future potential of AI in ICSI, including aiding in oocyte positioning by identifying the polar body location, sperm tracking and immobilisation prior to ICSI and automated sperm injection robots using AI algorithms. Another proposed use of AI was in the maintenance and monitoring of critical parameters and conditions in cryopreservation tanks, such as in the **TMRW** overwatch™ system, which includes an integrated algorithm for early prediction of future system failures.

### Embryo grading

**2.12.** A paper by Dimitriadis et al., 2022 gave a detailed review of the research in AI embryo, sperm and egg analysis to date, including a breakdown of analysis at different stages of embryo development. The study summarised that the use of AI in the fertility sector has so far focused on embryo assessment, but it has the potential for much wider application. Although AI is a promising tool, it is important to keep in mind that its ability to improve outcomes is yet to be proven in the literature.

**2.13.** Fordham et al., 2022 compared the embryo assessment of 39 embryologists with the performance of a deep neural network (DNN) on 136 time-lapse imaging videos of embryos that had reached the blastocysts stage. The average implantation prediction accuracy for the

embryologists was 51.9%, and the average accuracy of the embryologists when assessing top-quality and poor-quality embryos was 57.5% and 57.4%, respectively, and 44.6% for fair-quality embryos. Overall interobserver agreement was moderate, with the best agreement achieved in the poor- and top-quality group, while the agreement in the fair-quality group was lower. The DNN showed an overall accuracy rate of 62.5%, with accuracies of 62.2%, 61% and 65.6% for the poor, fair and top-quality groups, respectively. The AUC for the DNN was higher than that of the embryologists overall (0.70 DNN vs 0.61 embryologists) as well as in all of the Gardner groups (DNN vs embryologists-Poor: 0.69 vs 0.62; Fair: 0.67 vs 0.53; Top: 0.77 vs 0.54).

**2.14.** Berntsen et al., 2022 created an AI based embryo selection model, **iDAScore** v1.0 model, using 115,832 time-lapse images of embryos from 18 IVF centres. In an independent test set, the AI model sorted known implantation data (KID) embryos with an AUC of a receiver operating characteristic curve of 0.67 and all embryos with an AUC of 0.95. A clinic hold-out test showed that the model generalised to new clinics with an AUC range of 0.60–0.75 for KID embryos. Across different age subgroups, insemination method, incubation time, and transfer protocol, the AUC ranged between 0.63 and 0.69. The group summarised that the fully automated iDAScore v1.0 model was shown to perform at least as well as a manual embryo selection model. There were also suggestions that full automatisation of embryo scoring implies fewer manual evaluations and eliminates biases due to inter- and intra-observer variation. The effectiveness of iDAScore at predicting the live birth rate and miscarriage risk was confirmed in a retrospective study by Ueno et al., 2022.

**2.15.** Loewke et al., 2022 conducted a study to evaluate the benefit of an AI blastocyst ranking model for predicting clinical pregnancy using fetal heartbeat. The retrospective study used blastocyst images from 11 assisted reproductive technology centres in the United States of America. This included static images of 5,923 transferred blastocysts and 2,614 non-transferred blastocysts with aneuploid PGT-A results. The AUC of the AI model ranged from 0.6 to 0.7 and outperformed manual morphology grading overall and on a per-clinic basis. On visual inspection, the algorithm appeared to use similar features for classification as manual morphological grading. A secondary outcome of the study was to highlight potential limitations of AI assessment of embryos. The quality of the images captured, and focal plane of imaging used could impact the prediction of outcomes. Two potential sources of bias were also identified in microscope optics and the presence of holding micropipettes. The analysis of AI scores in relation to pregnancy rates showed that score differences of ≥0.1 (10%) correspond with improved pregnancy rates, whereas score differences of <0.1 may not be clinically meaningful.

**2.16.** A systematic review and data synthesis by Sfakianoudis et al., 2022 evaluated the predictive capabilities of AI based prediction models for IVF outcomes. The review found that many of the models were successful at accurately predicting outcomes, including clinical pregnancy, live birth, and ploidy status. The study also attempted to compare the performance of AI models and human grading. Although the studies included did not allow for a meta-analysis, the systematic review indicated that the AI-based prediction models perform similarly to the embryologists' evaluations. The group concluded that AI models appeared to be marginally more effective at the prediction of outcomes than embryologists, but there is some way to go before the models can surpass human performance.

### Prediction of ploidy status

**2.17.** AI has also been used to predict a patient's chance of creating a euploid, or aneuploid, embryo. Sun et al., 2022 used machine learning-based classifiers to analyse whole exome sequencing data and predict a patient's embryo aneuploidy risk. The study identified three candidate genes for aneuploidy risk that contribute the most to the predictive power of the model.

**2.18.** La Marca et al., 2022 performed a prospective analysis of 847 couples undergoing their first PGT-A cycle. After ovarian stimulation and oocyte insemination, 40.1% of couples had at least one blastocyst available for the PGT-A. Of 1068 blastocysts analysed, 33.6% were euploid. The study used machine learning models to determine the predictive potential of different co-variants on the blastocyst's euploid rate. Women's age and AMH with a positive association between the outcome, and AMH and a negative association between the outcome and female age appeared.

**2.19.** Chen et al., 2022 created a non-invasive preimplantation genetic testing for aneuploidy (niPGT-A) system using AI to analyse the chromosome sequencing of 345 paired blastocyst culture medium and whole blastocyst samples and predict blastocyst ploidy. The AI system was validated in 266 patients, and a blind prospective observational study was conducted to compare AI-guided niPGT-A with traditional niPGT-A analysis. Higher live birth rates and lower miscarriage rates were observed in A- and B-grade embryos versus C-grade embryos, and higher embryo utilisation rates through the AI system compared with traditional niPGT-A analysis.

**2.20.** Diakiw et al., 2022 used 5,050 static, day 5 images of embryos (from 2,438 women) with linked genetic data obtained from PGT-A to develop an AI model that predicted the embryo's ploidy status. Overall accuracy for the prediction of euploidy on a blind test dataset was 65.3%, with a sensitivity of 74.6%. When the blind test dataset was cleansed of poor quality and mislabeled images, overall accuracy increased to 77.4%. When using the genetics AI model to rank embryos in a cohort, the probability of the top-ranked embryo being euploid was 82.4%, which was 26.4% more effective than using random ranking, and between 13–19% more effective than using the Gardner score.

**2.21.** A study by Zou et al., 2022 created five machine learning models and two deep learning networks to predict the ploidy status of an embryo using 773 images of blastocysts undergoing PGT-A. They found that the predictive power of both ploidy prediction and implantation prediction was improved when they combined clinical features in the algorithms. However, the authors commented that the models for ploidy prediction were not highly predictive, suggesting they cannot replace preimplantation genetic testing currently.

## 3. Policy response

**3.1.** With the introduction of AI driven devices into the fertility sector, it is vital for the HFEA to consider its role as a regulator. For this reason, the HFEA has been monitoring the research and application of AI, as well as the wider policy responses being taken both within the UK and internationally.

**3.2.** There have been several key publications in the last year, including the UK's **National AI Strategy**, the Centre for Data Ethics and Innovation's **'Roadmap to an effective AI assurance eco-system'**, the Alan Turing Institute's **'Common regulatory capacity for AI'** paper and the Ada Lovelace Institute's **'Regulate to Innovate'** paper.

**3.3.** In June 2022, the MHRA published the response to their consultation on the **future regulation of medical devices in the UK**. This included chapter ten, 'Software as a medical device' (SaMD) (including AI as a medical device (AIaMD)). The consultation document included proposals to amend medical device regulation to ensure SaMD is regulated clearly, effectively and proportionally to the risks that the medical devices may present.

**3.4.** A definition of SaMD was proposed as *"a set of instructions that processes input data and creates output data".* Some respondents suggested that it was necessary to define AIaMD separately, but the consultation concluded that a specific definition of AIaMD would not be provided as this would risk being over prescriptive. However, they stated that there would be sufficient clarity of 'other terms' (which may include AIaMD) when producing supporting guidance.

**3.5.** The consultation laid out important considerations for SaMD, including SaMD being used in the UK but hosted in other jurisdictions, risk categorisation in line with the International Medical Device Regulators Forum (IMDRF) SaMD classification rule, the potential for an air lock classification rule for SaMD with a risk profile that is not well understood, pre- and post-market requirements and cyber security.

**3.6.** Section 65 considered AIaMD more specifically. The consultation concluded that there is little appetite for additional statutory changes for AIaMD regulation and that additional concerns will be addressed through robust guidance. There was overall support for the use of in vitro device regulation (IVDR)-type performance evaluation for diagnostic software, especially AI. However, the consultation response outlined that there was no intention of introducing mandatory logging of outputs for auditability.

**3.7.** In July 2022, the Office for Artificial Intelligence published their policy position paper "**Establishing a pro-innovation approach to regulating AI**". This paper indicated the government's position on the regulation of AI ahead of the publication of a wider white paper expected in late 2022.

**3.8.** The paper sets out the key challenges to AI regulation, including:

- **Lack of clarity** – Ambiguity of the UK's legal framework and application of regulatory bodies to AI because they have not been developed specifically with AI in mind. The extent to which UK laws apply to AI is often open to interpretation, making them hard to navigate.
- **Overlaps** – Laws and regulators' remits may regulate the same issue for the same reason leading to unnecessary, contradictory, or confusing layers of regulation.
- **Inconsistency** – Differences in the powers available to regulators to address the use of AI within their remits and the extent to which different regulators have started addressing these issues.
- **Gaps** – As UK legislation has not been developed with AI in mind, there may be current risks that are inadequately addressed and future risks that we may need to prepare for.

**3.9.** The paper does not provide a definition of AI as it aims to ensure that the system can capture future and current applications of AI. It instead proposed core characteristics of:

- **Adaptiveness 'explaining intent or logic'** – trained, once or continually, on data and executed according to patterns and connections which are not easily discernible to humans.
- **Autonomy 'assigning responsibility for action'** – automating complex cognitive tasks, where decisions can be made without express intent or ongoing control of a human.

**3.10.** The Office for AI also then outlines their proposed 'clear, innovation-friendly and flexible' approach for AI regulation:

- **Context-specific** - AI is a dynamic, general-purpose technology, and risks are dependent principally on the context of its application. Assessments of risk are to be made by the appropriate regulator.
- **Pro-innovation and risk-based** – Regulators are to focus on applications of AI that result in real, identifiable, unacceptable levels of risk. Regulators are to embed considerations of innovation, competition and proportionality in the implementation and enforcement of regulatory frameworks.
- **Coherent** –Set of cross-sectoral principles that regulators will develop into sector or domain-specific AI regulation measures. This includes:
  - Ensuring AI is used safely
  - Ensuring AI is technically secure, and functions as designed
  - Ensuring AI is appropriately transparent and explainable
  - Embed considerations of fairness into AI
  - Define legal persons' responsibility for AI governance
  - Clarify routes to redress or contestability
- **Proportionate and risk-based** – Implementation of cross-sectoral principles on a non-statutory basis which could be supplemented by clear guidance from the government. This would be kept under review and cannot rule out the need for future legislation.

**3.11.** The HFEA awaits the publication of the Office for AI's white paper this year to consider the direction of AI regulation within the UK.

**3.12.** However, the HFEA has also been taking actions to understand our role as a regulator in a sector that is increasingly using AI in its work. In order to prepare to undertake the necessary work to ensure the safety of patients within the regulated sector. This includes the creation of an internal working group for active discussion of the HFEA's role as a regulator, which has been running since 2021, considering existing guidelines and legislation that is applicable to the use of AI and collaborative conversations with other regulators, including the HTA, MHRA and CQC, to avoid the duplication of work and to identify regulator gaps and overlaps, as well as SCAAC retaining the issue as a priority for annual update and discussion.

## 4. Recommendations

**4.1.** Members are asked to:

- advise the Executive if they are aware of any other recent developments in AI relevant to fertility treatments or research.
- discuss their views of the impact AI will have on fertility treatment as technology advances, including practical and ethical challenges to their application.
- discuss their views on the Authority's regulatory interest around AI systems, scope, and limits, particularly considering the Office for AI's regulation of AI policy position paper.
- review whether any outputs from the HFEA are required, addressing the use, or regulation, of AI.

## References

Abdullah, K. A. L., Atazhanova, T., Chavez-Badiola, A., & Shivhare, S. B. (2022). Automation in ART: Paving the Way for the Future of Infertility Treatment. *Reproductive Sciences (Thousand Oaks, Calif.)*. https://doi.org/10.1007/S43032-022-00941-Y

Bardet, L., Excoffier, J. B., Salaun-Penquer, N., Ortala, M., Pasquier, M., Mathieu d'Argent, E., & Massin, N. (2022). Comparison of predictive models for cumulative live birth rate after treatment with ART. *Reproductive Biomedicine Online*, *45*(2), 246–255. https://doi.org/10.1016/J.RBMO.2022.03.020

Berntsen, J., Rimestad, J., Lassen, J. T., Tran, D., & Kragh, M. F. (2022). Robust and generalizable embryo selection based on artificial intelligence and time-lapse image sequences. *PloS One*, *17*(2). https://doi.org/10.1371/JOURNAL.PONE.0262661

C. N. Barreto, N., Castro, G. Z., Pereira, R. G., Pereira, F. A. N., Reis, F. M., Junior, W. M., Cavallo, I. K. D., & Gomes, K. B. (2022). Predicting in vitro fertilization success in the Brazilian public health system: a machine learning approach. *Medical & Biological Engineering & Computing*, *60*(7), 1851–1861. https://doi.org/10.1007/S11517-022-02569-1

Chen, L., Li, W., Liu, Y., Peng, Z., Cai, L., Zhang, N., Xu, J., Wang, L., Teng, X., Yao, Y., Zou, Y., Ma, M., Liu, J., Lu, S., Sun, H., & Yao, B. (2022). Non-invasive embryo selection strategy for clinical IVF to avoid wastage of potentially competent embryos. *Reproductive Biomedicine Online*, *45*(1), 26–34. https://doi.org/10.1016/J.RBMO.2022.03.006

Correa, N., Cerquides, J., Arcos, J. L., & Vassena, R. (2022). Supporting first FSH dosage for ovarian stimulation with machine learning. *Reproductive Biomedicine Online*. https://doi.org/10.1016/J.RBMO.2022.06.010

Diakiw, S. M., Hall, J. M. M., VerMilyea, M. D., Amin, J., Aizpurua, J., Giardini, L., Briones, Y. G., Lim, A. Y. X., Dakka, M. A., Nguyen, T. v., Perugini, D., & Perugini, M. (2022). Development of an artificial intelligence model for predicting the likelihood of human embryo euploidy based on blastocyst images from multiple imaging systems during IVF. *Human Reproduction (Oxford, England)*, *37*(8), 1746–1759. https://doi.org/10.1093/HUMREP/DEAC131

Dimitriadis, I., Zaninovic, N., Badiola, A. C., & Bormann, C. L. (2022). Artificial intelligence in the embryology laboratory: a review. *Reproductive Biomedicine Online*, *44*(3), 435–448. https://doi.org/10.1016/J.RBMO.2021.11.003

Fordham, D. E., Rosentraub, D., Polsky, A. L., Aviram, T., Wolf, Y., Perl, O., Devir, A., Rosentraub, S., Silver, D. H., Gold Zamir, Y., Bronstein, A. M., Lara Lara, M., ben Nagi, J., Alvarez, A., & Munné, S. (2022). Embryologist agreement when assessing blastocyst implantation probability: is data-driven prediction the solution to embryo assessment subjectivity? *Human Reproduction (Oxford, England)*. https://doi.org/10.1093/HUMREP/DEAC171

Hammer, K. C., Jiang, V. S., Kanakasabapathy, M. K., Thirumalaraju, P., Kandula, H., Dimitriadis, I., Souter, I., Bormann, C. L., & Shafiee, H. (2022). Using artificial intelligence to avoid human error in identifying embryos: a retrospective cohort study. *Journal of Assisted Reproduction and Genetics*. https://doi.org/10.1007/S10815-022-02585-Y

Khodabandelu, S., Basirat, Z., Khaleghi, S., Khafri, S., Montazery Kordy, H., & Golsorkhtabaramiri, M. (2022). Developing machine learning-based models to predict intrauterine insemination (IUI) success by address modeling challenges in imbalanced data and providing modification solutions for them. *BMC Medical Informatics and Decision Making*, *22*(1). https://doi.org/10.1186/S12911-022-01974-8

la Marca, A., Capuzzo, M., Longo, M., Imbrogno, M. G., Spedicato, G. A., Fiorentino, F., Spinella, F., Greco, P., Minasi, M. G., & Greco, E. (2022). The number and rate of euploid blastocysts in women

undergoing IVF/ICSI cycles are strongly dependent on ovarian reserve and female age. *Human Reproduction (Oxford, England)*. https://doi.org/10.1093/HUMREP/DEAC191

Letterie, G., MacDonald, A., & Shi, Z. (2022). An artificial intelligence platform to optimize workflow during ovarian stimulation and IVF: process improvement and outcome-based predictions. *Reproductive Biomedicine Online*, *44*(2), 254–260. https://doi.org/10.1016/J.RBMO.2021.10.006

Loewke, K., Cho, J. H., Brumar, C. D., Maeder-York, P., Barash, O., Malmsten, J. E., Zaninovic, N., Sakkas, D., Miller, K. A., Levy, M., & VerMilyea, M. D. (2022). Characterization of an artificial intelligence model for ranking static images of blastocyst stage embryos. *Fertility and Sterility*, *117*(3), 528–535. https://doi.org/10.1016/J.FERTNSTERT.2021.11.022

Mehrjerd, A., Rezaei, H., Eslami, S., & Khadem Ghaebi, N. (2022). Determination of Cut Off for Endometrial Thickness in Couples with Unexplained Infertility: Trustable AI. *Studies in Health Technology and Informatics*, *294*, 264–268. https://doi.org/10.3233/SHTI220450

Sfakianoudis, K., Maziotis, E., Grigoriadis, S., Pantou, A., Kokkini, G., Trypidi, A., Giannelou, P., Zikopoulos, A., Angeli, I., Vaxevanoglou, T., Pantos, K., & Simopoulou, M. (2022). Reporting on the Value of Artificial Intelligence in Predicting the Optimal Embryo for Transfer: A Systematic Review including Data Synthesis. *Biomedicines*, *10*(3). https://doi.org/10.3390/BIOMEDICINES10030697

Shen, L., Zhang, Y., Chen, W., & Yin, X. (2022). The Application of Artificial Intelligence in Predicting Embryo Transfer Outcome of Recurrent Implantation Failure. *Frontiers in Physiology*, *13*. https://doi.org/10.3389/FPHYS.2022.885661

Sun, S., Miller, M., Wang, Y., Tyc, K. M., Cao, X., Scott, R. T., Tao, X., Bromberg, Y., Schindler, K., & Xing, J. (2022). Predicting embryonic aneuploidy rate in IVF patients using whole-exome sequencing. *Human Genetics*. https://doi.org/10.1007/S00439-022-02450-Z

Ueno, S., Berntsen, J., Ito, M., Okimura, T., & Kato, K. (2022). Correlation between an annotation-free embryo scoring system based on deep learning and live birth/neonatal outcomes after single vitrified-warmed blastocyst transfer: a single-centre, large-cohort retrospective study. *Journal of Assisted Reproduction and Genetics*. https://doi.org/10.1007/s10815-022-02562-5

Wang, C. W., Kuo, C. Y., Chen, C. H., Hsieh, Y. H., & Su, E. C. Y. (2022). Predicting clinical pregnancy using clinical features and machine learning algorithms in in vitro fertilization. *PloS One*, *17*(6). https://doi.org/10.1371/JOURNAL.PONE.0267554

Wen, J. Y., Liu, C. F., Chung, M. T., & Tsai, Y. C. (2022). Artificial intelligence model to predict pregnancy and multiple pregnancy risk following in vitro fertilization-embryo transfer (IVF-ET). *Taiwanese Journal of Obstetrics and Gynecology*, *61*(5), 837–846. https://doi.org/10.1016/J.TJOG.2021.11.038

Zhang, Y., Shen, L., Yin, X., & Chen, W. (2022). Live-Birth Prediction of Natural-Cycle In Vitro Fertilization Using 57,558 Linked Cycle Records: A Machine Learning Perspective. *Frontiers in Endocrinology*, *13*. https://doi.org/10.3389/FENDO.2022.838087

Zou, Y., Pan, Y., Ge, N., Xu, Y., Gu, R., Li, Z., Fu, J., Gao, J., Sun, X., & Sun, Y. (2022). Can the combination of time-lapse parameters and clinical features predict embryonic ploidy status or implantation? *Reproductive Biomedicine Online*. https://doi.org/10.1016/J.RBMO.2022.06.007