Evidence base for fertility treatments

Strategic delivery:	☒ Safe, ethical, effective treatment	☐ Consistent outcomes and support	☐ Improving standards through intelligence
Details:			
Meeting	SCAAC		
Agenda item	7		
Paper number	HFEA (14/10/2019) 007		
Meeting date	14 October 2019		
Author	Dina Halai, Scientific Policy Manager		
Output:			
For information or recommendation?	For recommendation		
Recommendation	Members are asked to discuss: • Although RCTs are the only objective way to assess whether an intervention works, whether large data approaches can also be useful. • How much large data we need to make usable assumptions and account for confounding factors.		
Resource implications	None		
Implementation date	Recommendations will be considered by the HFEA for implementation in due course		
Communication(s)	Where necessary, website updates and wider engagement with the sector		
Organisational risk	☒ Low	☐ Medium	☐ High

1. Introduction

1.1. Evidence-based practice is when health professionals make a treatment decision with their patient, based on their clinical expertise, the preferences of the patient, and the best available evidence.

1.2. Finding the best evidence requires knowledge of the best quality, most appropriate sources, and how to use them. When reviewing the effectiveness of treatments, well-designed randomised controlled trials (RCTs) are thought to provide the most reliable source of evidence and therefore are considered to be the 'gold standard'. Unfortunately, there are many situations where good research studies have not yet been carried out, as outlined by Stocking et al, and this is particularly the case for so-called treatment add-ons in fertility treatment. The reasons for this are many and varied, including funding and the difficulty of sufficiently large sample sizes, but as things stands it is likely that many treatment add-ons will not have a well-designed RCT for the foreseeable future. The treatment add-ons consensus statement outlines that where there is no evidence to support safety and efficacy, treatment add-ons should only be offered to patients in a research setting. To this end, there may be a risk that treatments may not be able to go beyond the 'experimental' category.

1.3. The relative paucity of high-quality evidence raises questions for the HFEA's approach to providing patients and clinics with impartial evidence (through a 'traffic light' rating system) on the safety and effectiveness of treatment add-ons. At the event for person's responsible on fertility clinic licenses (PRs) in 2019 and the HFEA's Annual Conference 2019, some argued that the HFEA needs to consider if it should continue with an approach which uses RCTs as the key determinate of any assessment or if it should try to accommodate other types of evidence (notably retrospective studies of large data) into that assessment.

1.4. The issue is further complicated by a number of clinics beginning to report success with particular add-ons in their own clinic, and although such data is beginning to be shared at professional conferences it is not yet submitted for peer review. furthermore, a fertility clinic group informed us that they are unlikely to carry out RCTs because they hold a large amount of outcome data that they believe is a sufficient evidence source for them and it would not make sense for them to put patients in a control group[1] where they might better benefit from receiving the add-on.

1.5. Sticking with a traffic light rating based on RCTs ensures that our assessment is based on the highest quality studies but risks being overtaken by other publicly available research data; whilst accommodating data from other less robust sources risks diluting the objective quality of that assessment. At the June 2019 SCAAC meeting, members noted the current position regarding RCTs and agreed that a fuller discussion would be useful. This paper seeks SCAAC's views of the relative merits of these options.

2. Current assessment

2.1. A traffic light rating system was developed by the SCAAC in 2017 to give an indication of what conclusions could be read from the published evidence regarding the efficacy and safety of a

[1] In an RCT, subjects are randomly assigned to one of two groups: one (the experimental group) receiving the intervention that is being tested, and the other (the control group) receiving a conventional treatment.

selection of popular add-ons, but also to assess the quality of that evidence. The HFEA gives a green traffic light rating where there is more than one good quality RCT which shows that the procedure is effective and safe. We use an amber traffic light rating where there is a small or conflicting body of evidence, which means further research is still required and the add-on cannot be recommended for routine use. An add-on is red if there is no evidence to show that it is effective and safe. At present there are no add-ons rated as green.

2.2. The HFEA aims to support innovation and it is hoped that 'add-ons' will be a temporary category pending evidence on efficacy and safety for some add-ons, allowing them to be given a green light for use in routine clinical practice and therefore no longer being considered an add-on. Conversely, it is hoped that good quality research will also provide evidence where an add-on is not effective and/or safe for use in routine practice, and these will continue to exist as an add-on with a red traffic light rating to inform patients.

2.3. The HFEA identifies the published evidence on each add-on and annually seeks external advice from an expert in systematic reviews and evidence assessment to carry out independent assessments of the quality of evidence (using the GRADE methodology[2]) for each treatment add-on for the purpose of developing a traffic light rating.

2.4. In 2017, the evidence published in the last 10 years was sent to an independent reviewer. Where there was a large body of published evidence, only RCTs were sent in order to limit the time taken for the review. The reviewer then carried out an assessment of the quality of evidence for each add on using the GRADE methodology.

2.5. Expert advice informed that:

2.6. the methodological quality of studies should be reviewed without taking into account the biological or clinical plausibility of the treatment add-ons.

2.7. quality could be assessed by looking at PICO criteria (population, intervention, controls, outcomes) and factors including risk of bias, allocation concealment, selective outcome reporting and blinding.

2.8. The traffic light ratings are then considered by the SCAAC based on recommendations from the external expert assessor.

3. RCTs vs big data approaches

3.1. In health care, there are three types of knowledge:

- knowledge derived from research, sometimes called evidence;
- knowledge derived from audit and routinely collected data, sometimes called statistics;
- knowledge derived from the experience of patients/service users and professionals.

3.2. The approach to evidence should aim to ensure information:

- is balanced and reduces bias

[2] GRADE is an approach for grading the quality of evidence and the strength of recommendations. It was developed by the Grading of Recommendations, Assessment, Development and Evaluation Working Group.

- acknowledges uncertainty
- is produced using an explicit evidence-based process
- uses consistent language
- is impartial

3.3. RCTs use research methods aimed at reducing bias and that's why they are considered the only objective way to assess whether an intervention works. However, RCTs are difficult to carry out and require a large amount of time and money.

3.4. Macklon, et al. has proposed nonrandomized studies as an alternative to, and perhaps even an improvement over, RCTs in IVF. They argue that "*Advancements in health informatics present the opportunity to amass large amounts of detailed clinical data on the people undergoing IVF, the treatments they receive, and the outcomes of those treatments. These datasets can then be analysed with 'powerful algorithms' capable of 'exploiting confounders' in order to determine which treatments will work for individual patients."*

3.5. The arguments of Macklon et al. are not without controversy and Wilkinson et al are soon to publish an article in Human Reproduction evaluating the arguments against RCTs in IVF from a primarily methodological perspective. They argue that "*The abandonment of randomised evidence for algorithmic mining of large datasets will not improve our inferences. Statistics are not that capable, even if we rename them machine learning. RCTs are challenging, but rather than throw our hands in the air, we believe the answer is to focus efforts on how we can improve them."*

3.6. In the absence of good RCTs as evidence, an increasing proportion of the sector is relying on their own experiences and analysis of live birth rates and patient outcomes within their own clinics, to determine the suitability and effectiveness of treatments for patients. It is therefore essential to consider the appropriateness of alternative evidence in these circumstances.

4. Recommendations

4.1. Members are asked to discuss:

4.1.1. Although RCTs are the only objective way to assess whether an intervention works, whether large data approaches can also be useful.

4.1.2. How much large data we need to make usable assumptions and account for confounding factors.

4.2. Should SCAAC recommend broadening the range of data that the HFEA consider when assigning traffic light ratings to treatment add-ons, any changes will be considered by the HFEA for implementation in due course.

5. References

- Stocking K, Wilkinson J, Lensen S, Brison DR, Roberts SA, Vail A. Are interventions in reproductive medicine assessed for plausible and clinically relevant effects? A systematic review of power and precision in trials and meta-analyses. 2019 Feb. 34(4);659–665.
- Macklon NS, Ahuja KK, Fauser BCJM. Building an evidence base for IVF 'add-ons'. Reproductive BioMedicine Online. 2019 June; 38(6): 853 – 856.