Enacting evidence-based medicine in fertility care: tensions between commercialisation and knowledge standardisation

Journal:	*Sociology of Health & Illness*
Manuscript ID	SHI-00092-2021.R1
Manuscript Type:	Original Article
Subject Area:	Evidence-based medicine < RESEARCH AREAS, IVF < RESEARCH AREAS, Medical knowledge < RESEARCH AREAS, Medical practice/medical work < RESEARCH AREAS, STS (Science and technology studies) < RESEARCH AREAS, Ethnography < METHODS AND METHODOLOGY
Abstract:	In this article we explore the recent enactment of evidence-based medicine (EBM) in the field of fertility care. We aim to contribute to the medical sociology literature through an analysis of how evidence is produced, interpreted and institutionalised in a relatively new medical field such as in-vitro fertilisation (IVF), characterised by high uncertainty due to limited knowledge and high levels of commercialisation. Drawing on extensive ethnographic research conducted in England, this article explores the challenges IVF professionals encounter in producing credible data on the effectiveness of additional treatments, offering novel insights on the tensions between commercialisation and standardisation in the enactment of EBM. Extant medical sociology and Science and Technology Studies (STS) literature has shown the hidden professional work required to enact randomised control trials (RCT) in practice. Our analysis shows that this hidden work is not enough when there is a broader lack of standardisation in both clinical and research practices, as producing "good quality" evidence requires high levels of standardisation of knowledge production.

Enacting evidence-based medicine in fertility care: tensions between commercialisation and knowledge standardisation

Introduction

The rise of evidence-based medicine (EBM) in the last four decades has made it imperative to use "the best available evidence" in making decisions about the care of individual patients (Sackett et al., 1996). Hence, EMB has become a core subject of sociological analysis (for a review, see Timmermans, 2010a), which has explored its prevalence in a variety of medical fields, from alternative medicine to cancer services (Broom and Tovey, 2007; Trnka and Stöckelová, 2019). In particular, medical sociology has explored professional resistance to EBM (Jackson and Scambler, 2007; Traynor, 2009), the politics of evidence (Green, 2000; May, 2006) and how medical practice is legitimised and justified when, despite the demand for "good quality" evidence, professionals have to work in conditions of high uncertainty (Ducey and Nikoo, 2018).

Drawing on the medical sociology literature, in this article we explore the recent enactment of EBM in fertility care and IVF (in vitro fertilisation). Since its controversial introduction in 1978, IVF has evolved from an experimental research technique to an established clinical practice (see Franklin, 2013). The recent encounter of reproductive medicine with EBM makes this field a prime case for exploring the role of commercialisation in enacting EBM, as the sector is highly privatised (Van de Wiel, 2019), even in countries with strong national health systems (Gerrits, 2016).

Due to the recent emergence of EBM in this field, the extensive social science literature on fertility has overlooked this aspect. In contrast, the medical literature on IVF has highly debated this issue in the last few years (Harper et al., 2017; Wilkinson et al., 2019a; Repping, 2019). This literature focuses on the reasons for the lack of good quality evidence supporting the multitude of available additional treatments offered to couples, often at an extra cost, on top of their standard IVF cycles, which are known in the field as "add-ons".

The aim of this article is not to take sides in the debate on the availability and quality of the evidence supporting add-ons. Rather, it aims to close a gap in the medical sociology literature through an analysis of how evidence is produced, interpreted and institutionalised in a relatively new medical field such as IVF, characterised by high uncertainty due to limited knowledge and high levels of commercialisation. Drawing on extensive ethnographic research conducted in England, this article explores the challenges IVF professionals encounter in producing credible data on the effectiveness of add-ons, offering novel insights on the tensions between commercialisation and knowledge standardisation in the enactment of EBM.

The debate on evidence-based medicine in medical sociology

With the institutionalisation of EBM as the dominant model to assess the safety and efficacy of medical interventions in Western medicine, sociological analysis has focused on the processes of standardisation (Berg and Timmermans, 2000; Timmermans and Berg, 2003) and the social construction of evidence in medicine (Green, 2000). In the current EBM hierarchy of evidence, randomised control trials (RCTs, which are studies in which a number of similar

people are randomly assigned to two or more groups to test a specific drug or treatment) and their systematic reviews (that analyse data from different RCTs and compare their results) are are at the top of the pyramid. The dominant notion that systematic reviews of RCTs produce the only "good quality" evidence to assess medical interventions and guide public health recommendations has been sharply criticised in both the medical and sociological literature (Moreira, 2007; Timmermans and McKay, 2009; Will and Moreira, 2010; Maldonado Castañeda, 2019). Early sociological studies of EBM have shown that the classification of any other form of medical research (non-randomised, qualitative, observational and small scale) as producing "lower quality" evidence has fostered a systematic devaluation of other forms of relevant knowledge (Berg and Timmermans, 2000; Timmermans and Berg, 2003). Recent social science research has witnessed an increasing call for a more inclusive approach to evidence, noting that what is highly-rated in the dominant EBM model offers only extremely decontextualised information (Trnka and Stöckelová, 2019). However, the tenets of the EBM hierarchy of evidence have been institutionalised, including for fertility care (Authors, XXXX).

Although the biomedical community agrees rhetorically on what constitutes good quality evidence (i.e., systematic reviews of RCTs), not all the evidence produced through RCTs is automatically considered acceptable. Social studies of medical evidence have underlined the importance of studying the production and interpretation of evidence in practice. RCTs have built-in validity issues (Cartwright, 2007) and therefore are evaluated in professional practices in terms of appropriateness and credibility (Green, 2000), as results may be inadequate or misleading (Abraham, 2007). In addition, systematic reviews embed the values and interests of the groups that produce and engage with them (Moreira, 2007; Maldonado Castañeda, 2019) and policy decision-making is often subjective (Trnka and Stöckelová, 2019) and influenced by informal systems (Broom and Tovey, 2007).

In particular, the sociological literature on evidence in medicine stemming from science and technologies studies (STS) has explored the intricate and situated work required to conduct RCTs and produce credible data in various settings (see Will and Moreira, 2010), showing how research practices are just as complex as those of everyday clinical work. For instance, ethnographic accounts of RCTs show how the complexity and heterogeneity of conducting trials has to be silenced, requiring the enactment of additional professional practices such as handling research participants (Jespersen et al., 2014) and "cleaning" data to solidify trial results, i.e. repairing inconsistencies in the records without making data too "clean" to appear suspicious (Helgesson, 2010). These studies have also revealed the "hidden work" (Heaven, 2010) required to bridge the ontological divide between RCT researchers and clinicians in order to conduct successful trials. Overall, the STS literature shows that intensive professional work is required to enact the abstraction of trials into local research practices (Timmermans and Berg, 1997). This work includes various efforts to manage tensions at each site among different professional groups, interests and forms of expertise and can sometimes be a disruption for clinical practice (Timmermans, 2010b).

Background: the British Fertility Business and the resistance to EBM

As fertility care developed in the late 1970s and 1980s, at a time when EBM was not yet accepted as dominant practice, all IVF treatments currently considered standard were initially introduced without the support of clinical trials. Thereafter, the adoption of EBM has

encountered a strong resistance in the field (Cohen and Alikani, 2013; Dhont, 2013) and most RCTs included in systematic reviews are still classified as low quality (Wilkinson et al., 2019b). In particular, the lack of standardised outcome reporting in IVF is acknowledged in the field (Wilkinson et al., 2016).

In recent years, the question of evidence (and its scarcity) has become central in the public debate on fertility treatment in the UK, especially in relation to the proliferation of unproven treatment add-ons. The overall stagnation of IVF success rates has triggered an intense competition worldwide for the development of new tests and treatments that might improve pregnancy rates. However, several medical studies (Heneghan et al., 2016; Harper et al., 2017; Repping, 2019) denounce the lack of evidence for effectiveness that would support the use of these add-ons. The medical literature offers two main opposing explanations for the scarcity of evidence: some scholars suggest that this is mainly due to the commercialisation of the sector, as private clinics have an interest in continuing to sell expensive unproven treatments (Harper et al., 2017; Wilkinson et al., 2019a), while other scholars believe that, due to the specificity of fertility care, IVF treatments are not suitable to be assessed through EBM criteria (Cohen and Alikani, 2013; Dhont, 2013; Macklon et al., 2019).

The controversy has gained significant media attention in the UK, including a 2016 BBC Panorama documentary entitled "Inside Britain's Fertility Business" portraying private IVF clinics as taking financial advantage of vulnerable patients. Similarly, both in the medical and social science literature (Harper et al., 2017; Van de Wiel, 2019) the proliferation of add-ons is presented as a phenomenon that is due to the high-privatisation of the sector, regardless of the fact that several add-ons are widely used in NHS (public) clinics too (Authors, XXXX).

As a response to the heated public debate on add-ons, in 2017 the UK regulator of the sector, the Human Fertilisation and Embryology Authority (HFEA), launched a website where patients can find information on the effectiveness of the most widely-used add-ons (HFEA, 2019). Adopting a strict EBM approach, on this website add-ons are categorised using a "traffic light system": only add-ons with more than one high-quality RCT showing that the procedure is effective at improving live birth rates (LBR) are meant to be green (currently none); those with conflicting evidence are marked amber (currently five); and those with no evidence of safety or effectiveness are marked red (currently six). Addressing the need for responsible innovation, the HFEA introduced this tool to support informed choices for patients without denying them access to treatments.

In this article, we focus on the case of time-lapse monitoring (TLM), the most widely used add-on in the UK (Van de Wiel et al., 2020), which is marked amber in the HFEA traffic light system. TLM systems are incubators with integrated cameras able to monitor embryos while in the laboratory and optionally equipped with algorithms for embryo selection that are expected to improve IVF success rates. Interestingly, although there is no clear consensus over the ability of TLM to increase the number of babies born through IVF, these tools are unanimously considered safe and offering potential benefits by the professional community, which broadly legitimises TLM use in clinical practice (Authors, XXXX). TLM is a prime example of the current controversy on evidence in IVF, as despite the 11 currently available RCTs conducted in the last few years to assess its effectiveness, a recent systematic review (Armstrong et al., 2019) concluded that the evidence is low or very low quality overall. In addition to specific criticisms to some RCTs, the review shows a number of inconsistencies

across these studies that makes comparison impossible: heterogeneous patient populations, different laboratory practices (culture conditions, day of transfer and number of embryo transferred) and technical differences in the equipment, including various TLM systems and how they are used.

Methodology

To support our argument, we draw on ethnographic observations from five IVF clinics and 25 interviews with staff working at these clinics, as well as 18 interviews with key stakeholders. The data which inform our analysis are part of a bigger project looking at the introduction of TLM systems in IVF clinics in England. As they were categorised by the HFEA as "add-ons" in an early stage of the project, a significant portion of our data collection was aimed at understanding the production of evidence on the ground as well as professional views on evidence and clinical trials in IVF.

Ethnographic observations were carried out by both authors in NHS clinics in England where fertility treatment is provided, including the use of TLM tools and other add-ons. This involved observing laboratory and clinical practices (including gamete and embryo manipulation, egg collection and embryo transfer), as well as staff interactions in relation to the conduct of clinical trials, whenever this was required. In addition, researchers shadowed embryologists during their daily lab routines, use of TLM tools and interactions with other clinic staff. Clinics were selected based on their use of add-ons, availability and willingness to participate in the study. Further details about clinics cannot be disclosed due to reasons of confidentiality. Clinics agreed to participate in the research study and all staff were informed ahead of time about study procedures. Each person observed signed a consent form prior to the start of our research. In addition to university ethics approval, we received research clearance from the Health Research Authority and each clinic site separately. Observations took place in five clinics, for a total of 230 hours. Dependent on clinic availability, total observation time for each clinic ranged between 25 and 80 hours, with a minimum of three days of observation time for each. Researchers took notes as they observed staff.

We paid special attention to the clinics' procedures as they related to their participation in RCTs for add-ons. All five clinics observed were involved in at least one IVF RCT. However, the level of recruitment differed. Some sites had occasional patients that opted into trials, while others had higher opt-in numbers. This varied considerably due to clinic size, recruitment targets and staff involvement. However, we were able to witness the trial protocol being implemented at least once in all sites. We also talked to staff more widely about trial practices during observation and interviews.

Professionals working in the observation sites were personally approached by the researchers regarding interviews. Participation was voluntary and interviewees signed an additional consent form. We conducted a total of 25 interviews with NHS staff, including embryologists, consultants, nurses and clinic directors. We also include data from 18 key stakeholder interviews (see Figure 1 below for interview details). These stakeholders were recruited via email based on their IVF and add-on expertise and included clinicians, researchers and representatives of professional bodies. All interviews included in the analysis were conducted by the authors between June 2017 and July 2019, either face-to-face (35 interviews) or over Skype or phone (eight interviews) where meeting in person was not feasible. These

lasted between 35 and 100 minutes and were recorded and transcribed. Our interview guide focused on obtaining participants' experiences with using add-ons, conducting clinical trials, opinions on new and controversial add-ons as well as issues with producing reliable evidence in the IVF sector. We aimed to collect responses without expressing any personal views on evidence and EBM practice in IVF.

[Figure 1 here]

We entered interview and observation data into NVivo, a software package used for the analysis of qualitative data. The data were coded in two ways: 1) deductively, by looking for themes found in the previous literature or prior discussions between authors and 2) inductively, where we aimed to identify new codes, specific to our case study. When new codes emerged, these were then discussed between the authors through a continuous process until we reached agreement on all codes as well as the larger themes of the research. Thus, the analysis process was continual and open-ended. This was consistent with Grounded Theory methods and principles. This approach was used to guide the data analysis process, from developing smaller codes to integrating them into the larger themes discussed in this article (Glaser and Strauss, 2017). When coding, we paid close attention to all instances where trials, add-ons and evidence are discussed or appear in our observation notes. Through a process of continuous coding and discussion, we made connections between data on add-ons, trial practices, successes, difficulties, respondents' views on evidence and its integration in IVF, discussion of evidence and its importance, as well as arguments for and against using strict EBM protocols in IVF. Through the process of coding, analysis and relating our data to the previous social science literature on medical evidence, we were able to organise our findings by focusing on the two opposing explanations for the lack of good quality evidence in this field suggested by the medical literature: the commercialisation of the provision of fertility treatment and the misalignment between IVF clinical practice and the criteria of RCT protocols. Within each section, we grouped specific challenges to evidence production that respondents outlined in interviews or that were experienced during our observations and positioned our analysis in the broader social science literature on IVF.

The effects of the commercialisation of the IVF sector on evidence production

Focusing on the challenges to evidence production as introduced above, our findings are presented here in two parts. In this first section, we analyse how the commercialisation of IVF shapes laboratory practices and hinders the participation in research of both clinics and patients; while the next section explores the challenges posed by some characteristic features of IVF to the implementation of RCT protocols.

As we have outlined elsewhere in more detail (Authors, XXXX), the pace of innovation in IVF is incredibly fast. One of the effects of the commercialisation of the fertility sector is the availability of a wide range of products and technologies for the embryology team to choose from, which introduces a myriad of potential variations into local laboratory practices. For example, different clinics use different media in various types of embryo petri dishes, different types of incubators and different algorithms with their TLM machines. An embryologist highlighted the issue of conducting research despite heterogeneous practices in fertility clinics:

We use our in-house algorithm which is simple, is based on speed of development and morphological grade, you know. And we also look at early cleavage which, again, there's debates in the literature whether early cleavage makes any difference, but we've consistently found that the embryos that we select for early cleavage and transfer are more vulnerable, but that's in our practice, in our lab using our culture media. And this is one of the problems when you're looking at embryology as a whole is that everybody does things differently and so even when you have a multi-centre trial you're going to have heterogeneity in practice within the multi-centre trial quite often and that can be a huge confounder for the trial (Embryologist, Professional body representative).

The interviewee highlights how the way professionals practise embryology is contingent upon local variables. Clinics have to decide for themselves which parameters to include in their TLM algorithms, in addition to picking the embryo media that they think is best suited to their needs. As the medical literature on embryo selection is still emerging and conflicting, these decisions are highly shaped by local circumstances. During interviews, our respondents stressed repeatedly that comparisons across clinics are difficult due to differing lab routines.

In addition, due to the highly competitive market of fertility services, clinics often feel under pressure to adopt cutting-edge treatments and technologies, developing differing approaches to their use. Although these treatments and technologies are offered before the emergence of conclusive efficacy evidence, in some clinics they become part of the local standard practices through internal validation. This situation can produce conflict between local lab practices and RCT protocols, as a professional involved in conducting a clinical trial on TLM explains:

A lot of places [clinics] *have decided that the time lapse is more effective than standard care so they don't feel like they can randomise someone's standard of care when they truly believe in time lapse. So yeah, they need a machine, they need to have equipment, they need support from the embryologists as well because this is an embryology led trial, obviously they do the randomisation, they do the intervention. So we've had a few* [clinics] *where the clinicians are really for it and then you actually get to speak to the embryologists and then all of a sudden they're not as keen* (Clinical research professional, RCT coordinator).

Given that some clinics have already adopted TLM as standard practice, randomisation with standard incubators might be seen as taking a step back and not offering patients an add-on that is already included in everyone's treatment. During our lab observations, we witnessed some organisational resistance to the idea that adopted technologies might later prove to be ineffective. Once TLM is introduced in a lab, professionals hope that it is only a matter of time before research confirms that this was the right choice. Trials for add-ons are hard to support without the full cooperation of embryologists, who are at the forefront of administering the trial intervention (in this case, TLM incubators). Enthusiastic participation from staff is needed to persuade the IVF centre to participate, especially if the trial is not privately funded, in which case there is little financial benefit for a busy clinic to conduct RCTs. As a result, even minor disruptions to clinic practice might turn into a key factor in deciding whether or not to participate. During observation in one of the clinics, for instance, after a long negotiation with a potential new RCT site, its participation was withdrawn due to a technical issue (a different gas mixture in their incubators): as this would have required further changes to their equipment and local practices, staff at the centre were unwilling to participate.

The privatisation of the UK IVF sector was seen as negatively affecting not only clinic but also patient willingness to participate in trials. As age is a significant predictor of IVF success, patients' sense of urgency about getting pregnant, which is well known in the literature (Franklin, 1997; Thompson, 2005), represents an hinderance to their participation in trials. In addition, as the provision of fertility care is highly privatised, often patients come into their clinic with certain ideas about the treatments that they want to have. When considering infertility treatment and choosing a clinic, individuals might pick up pamphlets from companies or private clinics that advertise new and attractive treatments. One can also read about new IVF add-ons simply by performing a quick Google search. This poses a challenge for RCTs, which by design do not guarantee that the patient will actually get randomised into the study arm that provides the desired treatment. For example, during observations professionals who have had experience with an endometrial scratch RCT told us that, upon hearing they might go into the control arm of the trial (and consequently not get this particular add-on), some patients asked to pay for the treatment and try it out, independent of the trial being conducted. During observations, practitioners also talked about the disappointment that IVF patients experience when they find out that trial participation does not always mean they get the intervention. One of the NHS research nurses we interviewed explained that something she *"always focuses on with patients"* is *"to make sure that they're fully informed"* that they *"might not necessarily get the treatment because that's research: you might get the placebo or whatever the study is looking into."* This, coupled with the luck of public funding to cover IVF cycles (some patients wait for months and sometimes years before receiving referrals and clinic consultations), poses significant challenges to patient recruitment.

Patients who do participate are often the ones that genuinely want to contribute to the IVF knowledge base. However, in the incredibly financially and emotionally stressful context of infertility treatment (Franklin, 1997) and the stigma attached to it (Inhorn, 2003), many feel that participating in research at such a difficult time is too much of a burden. A professional commented on the issue of patient recruitment:

It might be that they [patients] *are then required to come in for another appointment, that's difficult. It might be they just don't want to, they just want to get on with their treatment. And just stick to what there is there, they don't want nothing new. There's probably lots of reasons. And then the trials themselves can be so complex, can't they, to set up, to make it so that at the end of the trial you've got an answer* (Fertility nurse, NHS clinic).

The problems with patient recruitment directly impacts the ability of RCTs to reach the required size for findings to be statistically significant. In addition, participants were also very conscious of issues surrounding the privatisation of the UK IVF sector and the consequences this has on the availability of reliable outcome data, due to bigger structural and commercial issues that have impeded the collection of large data sets. For instance, some respondents suggested that companies must have more data which is not always reported, and argued that countries where IVF has strong state support and funding are able to collect much more reliable data on new treatments.

Standardisation and protocol challenges in fertility care

In this section, we discuss the second potential explanation suggested by the IVF medical literature for the lack of good quality evidence in this field, by focusing on the efforts made by professionals in the attempt to align certain characteristic features of IVF with the standardising logics of RCT protocols.

The two main essential criteria of RCT protocols are randomisation and blinding. Randomisation is the process of assigning patients by chance to groups that receive different treatments (for instance, a new treatment to be compared with standard therapy). EBM best practice aims for the blinding of both patient and care provider in RCTs, so that neither know which treatment is being administered.

Observing how randomisation is performed for a TLM trial with three arms (no intervention, TLM use, and TLM use with algorithm), we witnessed some of the challenges posed by integrating the trial protocol into the daily practices of the lab. As we wrote in trial observation notes:

"*According to the protocol, randomisation should be done only subject to availability of a space in incubators for each arm of the study* [and this] *had an influence on the decision of what kind of care is offered to the other patients*" (author fieldnote).

Aligning the availability of space in each incubator while following the RCT protocol required some coordination: working staff had to ensure that non-trial patients had access to the necessary incubators, while at the same time they had to make sure that trial embryos were put in a specific incubator determined by the randomisation process. There was often little preparation that could be done in advance because it was only after embryos developed that staff knew if a patient had enough available embryos (the protocol required a minimum of 3) and could therefore participate in the trial. While the burden to ensure best care for all patients was on embryologists conducting the trial, they did not participate in the decision regarding the criteria for randomisation. Interestingly, in this trial randomisation did not include any information regarding the quality of gametes or embryos, but only general information about the patients (such as age, participation in other trials, and number of IVF cycles). Therefore, filling in the randomisation form was seen as a tedious and boring housekeeping task that disrupted the flow of work in the lab and was usually assigned to more junior embryologists.

Guaranteeing double blinding (so that neither patients nor professionals knew which treatment was being administered) proved to be even harder. During observations and interviews, our participants noted that add-ons, many of which are technological interventions or procedures, cannot be trialled in the same way as a drug where double blinding is more feasible. For example, TLM, as an embryology tool, does not allow for the blinding of embryologists who are tasked with physically placing the embryos in their respective incubators. One interviewee explains:

The only way you can blind, you could blind it to the patients because they wouldn't know which incubator, you could not tell them what it's gone in, but from an embryology point of view there is absolutely no way you could blind whether it's in standard or in a time lapse because it's obvious which incubator it's in. So you wouldn't be able to do that (Embryology lab director, NHS clinic).

To minimise the risk of sample mismatching, IVF labs are legally required to label all labware and implement witnessing protocols so that embryologists can be constantly able to identify reproductive samples. Therefore, in the case of TLM tools, as in all interventions on gametes or embryos, it is only the patient and the consultant who can be blinded. This feature of IVF is limiting for best practice protocols, yet as a clinician/researcher explained to us, trials have to adapt to the reality on the ground:

There will always be that risk that not everyone is blinded to the intervention, as with a study looking at medication, for example, where no one actually knows what you're taking, where you can blind everything. I think that is, that is a major flaw in IVF studies that we're never going to be able to properly iron out. But we can, you know, it doesn't mean that we shouldn't be doing studies. I just think that it's a bias that we're just going to have to live with. We've just got to recognise it and take it into account. Because when you unblind people, there's always that temptation to reveal that to the patient or to alter your practice as a result of being unblinded, so unfortunately I think it, even though you think it shouldn't matter I think sometimes it does matter being unblinded as an embryologist (Clinical research professional, author of a systematic review).

During observations professionals were concerned about the possibility of consultants accidentally walking into the lab and finding out which arm of the trial the patient is in. This is especially the case in smaller clinics where patients are well known and remembered by clinic staff. In one of the clinics, posters were used to signal to other staff not to approach a lab area where they could become unblinded/see the patient's course of treatment. It is thus obvious that blinding in IVF trials is not only unfeasible in many cases, but it is also a practice that might require ongoing work from staff for RCT standards to be upheld.

Another area of contention between the standardising logics of EBM protocols and fertility practices that participants highlighted during our study is how RCT outcomes are measured. According to protocols that follow RCT gold standards, the outcome that is considered the best indication of efficiency in IVF trials is the live birth rate (LBR), which implies that a pregnancy has been carried to term and resulted in a healthy baby. However, some clinicians complained that many RCTs, especially those funded by private companies, only look at pregnancy rates. This is considered to be a significant shortcoming, given that not all pregnancies are carried to term and miscarriages occur fairly often in patients with infertility issues. A researcher/clinician complained about the lack of standardisation in IVF outcome reporting:

It makes it really difficult to assess outcomes. And then they also have different end points. Some say pregnancy is when they have a positive pregnancy test, whereas other studies say, you know, it's when you see a heartbeat on a scan or … It makes it really difficult to bring together data from all these studies. And you see hardly any fertility studies report live births. You'd think that that would be the number one thing people are interested in, you know, patients … I would want to know as a patient how likely am I to come out of this with a child. But, interestingly, that's often not reported. And again, with stillbirth it's rare but at least miscarriage should be reported and often it isn't. So there should be a move to standardise outcome reporting in fertility trials (Fertility consultant, RCT lead investigator).

RCT outcome reporting seems to suffer from a similar confusing and even misleading representation as IVF "success rates" (see Inhorn, 2003; Gerrits, 2016), especially when clinics are commercial institutions. However, when trial outcomes are not reported uniformly, it also

becomes significantly more difficult to conduct systematic reviews of RCTs, hindering the production of good quality evidence. To overcome this issue, a majority of those interviewed and observed performing trial duties in the clinic expressed a wish for more standardisation in IVF clinical trials. Although LBRs are, for most, the desirable outcome to be reported in RCTs, there are also voices arguing for a larger re-evaluation of what is being assessed. For example, in the case of TLM, which is essentially an incubator meant to help with the assessment and selection of embryos, there is no evaluation of other benefits that it might offer patients, such as shortening the time to pregnancy. The lab tool can also offer valuable information to patients about their embryos, thus shaping clinical practice (for a more in-depth analysis about the functions of TLM see Authors, XXXX). The critique of outcomes being assessed in a TLM RCT was summarised by an embryologist who values the tool for its lab function:

So the way that I see it is that this is not a drug, this is not something that on one arm you have the drug, on the other arm you don't have the drug. I think that's the wrong way to see it. […] If an embryo is pink and another one is blue, you know, is there a difference in their implantation rate? As opposed to should we blind us to what is the colour of the embryo… And maybe in that arm that didn't use the technology they were all blue which is better than pink. So it's a diagnostic tool and it should be assessed as a diagnostic tool, not as a drug intervention. So this is my personal opinion on the matter. […] I feel I understand the technology and I feel it's being misinterpreted in the literature. (Embryology lab director, Private clinic)

The embryologist highlights that the quality of the embryos is more important than whether or not TLM was used to select them. As such, she sees TLM as a diagnostic tool that, as information is gathered through it, should be able to assess embryos more accurately regarding their implantation potential. Such opinions raise larger questions about each IVF intervention and whether evaluation criteria might or should look different for different treatments, also given that there are only few add-ons that are similar to a drug intervention.

Discussion: Tensions between commercialisation and standardisation in IVF knowledge production

In this article, we analysed the challenges of evidence production in IVF, by exploring the effects of the commercialisation of the sector and protocol challenges in fertility care. Our analysis shows that, although both these two issues are relevant to explain the complexity of knowledge production in this sector, they are not as mutually exclusive or polarised as the current medical literature suggests, but are rather core elements of a broader set of complex clinic practices.

First, the privatisation of the sector emerges as a central question in our findings, but with different nuances compared to how this is presented in the literature and public debate. IVF professionals (both working in NHS and private clinics) identify the lack of funding for IVF treatments as negatively affecting clinic and patient willingness to participate in RCTs. Due to the commercialisation of the field, the availability of a wide range of products and technologies for embryology laboratories has fostered a huge variation in IVF practices. In addition, as new IVF treatments are promptly adopted, their implementation in clinics can vary significantly, as our findings on the use of TLM and algorithms illustrate in detail. All these variations among labs make adherence to RCT protocols difficult to enact, as we discuss in greater detail in what

follows, and therefore reduce the willingness of clinics to participate in trials. The high privatisation of the sector also has a significant impact on patient recruitment and therefore the ability of RCTs to reach the required size for findings to be statistically significant in a timeframe compatible with the current research practices. Due to the lack of funding, the waiting list for NHS-funded IVF treatments can be significant, when funding is available at all. Treatments in private clinics are the only option for many couples. As age is a significant predictor of IVF success, patients manifest a sense of urgency about getting pregnant (Franklin, 1997; Thompson, 2005) that hinders their willingness to participate in trials, especially as these do not guarantee them access to a specific add-on. In addition, in the incredibly stressful context of infertility treatment the additional commitment required to participate in a trial can appear as an unnecessary burden for patients, especially considering that due to age restriction and time needed to produce RCT results, these patients will not benefit from them personally. Our findings do not suggest that labs resist research per se or avoid proving the ineffectiveness of treatments, as suggested by critics. Rather, the case is that heterogeneous clinic practices and the rapid adoption of innovations do not facilitate a seamless application of RCT protocols developed with standardisation in mind. This duality creates friction between stakeholders and makes the participation in trials less appealing for both clinics and patients.

Second, our analysis shows that some requirements of the RCT standard protocol such as blinding and randomisation pose challenges to the implementation of trials in fertility care, as suggested in the literature (Cohen and Alikani, 2013; Dhont, 2013; Macklon et al., 2019). For instance, in our findings we illustrate how blinding in IVF trials is sometimes unfeasible following current standard RCT procedures. However, the STS literature exploring RCTs in other medical fields has already shown that enacting the abstractions of an RCT protocol (Timmermans and Berg, 1997) does require intensive professional work to be integrated into the daily clinical practices in order to produce reliable data (Heaven, 2010; Helgesson, 2010; Jespersen et al., 2014). We argue that adjusting RCT protocols to IVF practices requires not just this intensive work, but rather a process of institutionalisation of knowledge production that is a prerequisite for producing credible evidence. Our findings show that the process of institutionalisation of knowledge production is still ongoing in this field and conflicting practices and epistemologies are simultaneously active and supported by different groups.

Overall, what emerges from our analysis is that the two critiques emerging in the debate are actually shifting attention away from the two main challenges of IVF evidence production that we identify as a lack of standardisation of both IVF practices and IVF RCT outcomes.

As we mentioned above, the heterogeneity of local laboratory practices in the fertility sector not only hinders the ability of individual labs to participate in multi-centre trials, but also weakens the evidence that might be produced in the trial even in the case of a successful participation. We claim that the adherence to RCT protocols does not only disrupt clinical practices as previously showed in the literature (Timmermans, 2010b), but that the lack of standardisation of clinic practices makes the standardisation of research practices unfeasible in the current conditions. Focusing on the example of TLM, the lack of standardisation does not pertain only to the high variations in its uses (i.e., as an incubator only or with the support of a variety of locally built selection algorithms often developed on a clinic's own data and patient population), but also the conceptualisation of the tool itself (i.e., as an add-on whose efficiency has to be proved or a better piece of equipment – regardless of its ability to improve IVF results – that should be adopted as standard practice).

In addition, our analysis shows how the lack of standardisation of relevant outcomes of RCTs illustrate the ontological divide (Heaven, 2010) between RCT researchers and IVF practitioners – and even among IVF practitioners themselves. We claim that the field is characterised not only by highly heterogeneous laboratory practices but also by conflicting epistemologies that are supported by different professional groups. For instance, looking at the case of TLM there are opposing perspectives on what the relevant outcomes of trials should be. RCT researchers argue that LBR is the only acceptable measure and other measures (such as pregnancy rates) are inadequate due to the high chances for fertility patients of having a miscarriage. While this position is rhetorically shared by some IVF professionals, others question the validity of LBR as a measure of the effectiveness of TLM, as these are just diagnostic tools (as many other interventions) unable to change the biological materials they are monitoring. Therefore, they argue for a larger re-evaluation of what is being assessed and of the other benefits that some interventions might offer patients (in the case of TLM, for instance, shortening the time to pregnancy or offering valuable information to patients about their embryos). Similarly, some of our participants ascribe the lack of consistency in RCT results to a desire to manipulate data, while others challenge the expectations of trial researchers on how results should be recorded. We argue that, in the absence of standardised processes of knowledge production, the current debate over the most appropriate manner to standardise RCT result collection in this field is genuinely due to conflicting epistemologies between the logic of trial researchers and that of IVF professionals.

Conclusion

This article aims to contribute to the critical social science debate on EMB, by offering an ethnographic account of the enactment of RCTs in the field of fertility care. Extant medical sociology and STS literature has shown the hidden professional work required to enact RCTs in practice. Our analysis shows that this hidden work is not enough when there is a broader lack of standardisation in both clinical and research practices. We argue that the tensions between commercialisation and knowledge standardisation identified in this article further explain the challenges medical professionals face to substantially adhere to EBM tenets, as producing "good quality" evidence requires high levels of standardisation of knowledge production. Finally, we acknowledge that our case focuses on a highly privatised but largely legitimised tool in embryology. To further explore the role of commercialisation in the adoption of EBM, further research is required to explore more problematic, risky or unsafe medical interventions.

References

Abraham, J. (2007). Drug trials and evidence bases in international regulatory context, *Biosocieties*. 2(1), 41-57.

Armstrong, S., Bhide, P., Jordan, V., Pacey, A., Marjoribanks, J., & Farquhar, C. (2019). Time-lapse systems for embryo incubation and assessment in assisted reproduction, *Cochrane Database of Systematic Reviews*, 5.

Berg, M., & Timmermans, S. (2000). Orders and their others: On the constitution of universalities in medical work. *Configurations,* 8(1), 3, 1-6.

Broom, A., & Tovey, P. (2007). Therapeutic pluralism? Evidence, power and legitimacy in UK cancer services. *Sociology of Health & Illness*, 29(4), 551-569.

Cartwright, N. (2007). Are RCTs the gold standard?. *BioSocieties*, *2*(1), 11-20.

Cohen, J., & Alikani, M. (2013) Evidence-based medicine and its application in clinical preimplantation embryology. *Reproductive Biomed Online*, 27, 547-561.

Dhont, M. (2013). Evidence-based reproductive medicine: a critical appraisal. *Facts, views & vision in ObGyn*, *5*(3), 233.

Ducev, A., & Nikoo, S. (2018). Formats of responsibility: elective surgery in the era of evidence-based medicine. *Sociology of Health & Illness*, 40(3), 494-507.

Franklin, S. (1997). *Embodied progress: A cultural account of assisted conception*. Routledge, London.

Franklin, S. (2013). *Biological relatives: IVF, stem cells and the future of kinship*. Duke University Press, Durham.

Gerrits, T. (2016). *Patient-centred IVF: bioethics and care in a Dutch clinic*. Berghahn books, New York and Oxford.

Glaser, B. G., & Strauss, A. L. (2017). *Discovery of grounded theory: Strategies for qualitative research*. Routledge, London.

Green, J. (2000). Epistemology, evidence and experience: evidence based health care in the work of Accident Alliances. *Sociology of Health & Illness*, 22(4), 453-476.

Harper, J., Jackson, E., Sermon, K., Aitken, R. J., Harbottle, S., Mocanu, E., ... & Lundin, K. (2017). Adjuncts in the IVF laboratory: where is the evidence for 'add-on'interventions?. *Human Reproduction*, 32(3), 485-491.

Heaven, B. (2010). Bridging the ontological divide: different social worlds in the conduct of a pilot study. In Will, C., & Moreira, T., (2010). *Medical Proofs, Social Experiments: Clinical Trials in Shifting Contexts*. Ashgate, Farnham (UK) and Burlington (US), 33-48.

Helgesson, C. F. (2010). From dirty data to credible scientific evidence: some practices used to clean data in large randomised clinical trials. In Will, C., & Moreira, T., (2010). *Medical Proofs, Social Experiments: Clinical Trials in Shifting Contexts*. Ashgate, Farnham (UK) and Burlington (US), 49-64.

Heneghan, C., Spencer, E. A., Bobrovitz, N., Collins, D. R. J., Nunan, D., Plüddemann, A., ... & Tompson, A. (2016). Lack of evidence for interventions offered in UK fertility centres. *British Medical Journal*, 355, i6295.

Human Fertility and Embryology Authority. (2019). Treatment add-ons. Retrieved at https://www.hfea.gov.uk/treatments/explore-all-treatments/treatment-add-ons/

Human Fertility and Embryology Authority (2020) Fertility treatment 2018: Trends and figures. Retrieved at https://www.hfea.gov.uk/about-us/publications/research-and-data/fertility-treatment-2018-trends-and-figures/

Inhorn, M. C. (2003). *Local babies, global science: gender, religion and in vitro fertilization in Egypt*. Routledge, London.

Jackson, S., & Scambler, G. (2007). Perceptions of evidence-based medicine: traditional acupuncturists in the UK and resistance to biomedical modes of evaluation. *Sociology of Health & Illness*, 29(3), 412-429.

Jespersen, A. P., Bønnelycke, J., & Eriksen, H. H. (2014). Careful science? Bodywork and care practices in randomised clinical trials. *Sociology of Health & Illness*, *36*(5), 655-669.

Macklon, N. S., Ahuja, K. K., & Fauser, B. C. J. M. (2019). Building an evidence base for IVF 'add-ons'. *Reproductive biomedicine online*, 38(6), 853-856.

Maldonado Castañeda, O.J. (2019). (Un)contested evidence: scientific literature, systematic reviews and the politics of evidence in the introduction of HPV vaccines in Colombia. *Sociology of Health & Illness*, 41, 81-94.

May, C. (2006). Mobilising modern facts: health technology assessment and the politics of evidence. *Sociology of Health & Illness*, 28(5), 513-532.

Moreira, T. (2007). Entangled evidence: knowledge making in systematic reviews in healthcare. *Sociology of Health & Illness*, 29(2), 180-197.

Repping, S. (2019). Evidence-based medicine and infertility treatment. *The Lancet*, 393(10170), 380-382.

Sackett, D.L., Rosenberg, W.M.C., Gray, J.A.M., Haynes, R.B. and Richardson, W.S. (1996) Evidence based medicine: what it is and what it isn't, *British Medical Journal*, 312, 71–2.

Timmermans, S. (2010a). Evidence-based medicine: Sociological explorations. In Bird, C. E., Conrad, P., Fremont, A. M., & Timmermans, S. (Eds.). *Handbook of Medical Sociology*. Vanderbilt University Pres, 307-323.

Timmermans, S. (2010b). Reconciling research with medical care in RCTs. In Will, C., & Moreira, T., *Medical Proofs, Social Experiments: Clinical Trials in Shifting Contexts*. Ashgate, Farnham (UK) and Burlington (US), 17-31.

Timmermans, S., & Berg, M. (1997). Standardization in action: achieving local universality through medical protocols. *Social Studies of Science*, 27(2), 273-305.

Timmermans, S., & Berg, M. (2003). *The gold standard: The challenge of evidence-based medicine and standardization in health care*. Temple University Press, Philadelphia, PA.

Timmermans, S., & McKay, T. (2009). Clinical trials as treatment option: bioethics and health care disparities in substance dependency, *Social Science & Medicine*, 69(12), 1784-1790.

Thompson, C. (2005). *Making parents: The ontological choreography of reproductive technologies*. MIT press, Cambridge, MA.

Traynor, M. (2009). Indeterminacy and technicality revisited: how medicine and nursing have responded to the evidence based movement. *Sociology of Health & Illness*, 31(4), 494-507.

Trnka, S., & Stöckelová, T. (2019). Equality, efficiency and effectiveness: going beyond RCTs in AL Cochrane's vision of health care. *Sociology of Health & Illness*, 41(2), 234-248.

Van de Wiel, L. (2019). The datafication of reproduction: Time-lapse embryo imaging and the commercialisation of IVF. *Sociology of Health & Illness*, 41(51), 193-209.

Van de Wiel, L., Wilkinson, J., Athanasiou, P. and Harper, J. (2020). The prevalence, promotion and pricing of three IVF add-ons on fertility clinic websites. *Reproductive BioMedicine Online*, 41(5), 801-806.

Wilkinson, J., Roberts, S.A., Showell, M., Brison, D.R., & Vail, A., (2016). No common denominator: a review of outcome measures in IVF RCTs, *Human Reproduction*, 31(12), 2714-2722.

Wilkinson, J., Bhattacharya, S., Duffy, J. M. N., Kamath, M. S., Marjoribanks, J., Repping, S., ... & Farquhar, C. M. (2019a). Reproductive medicine: still more ART than science?. *BJOG: An International Journal of Obstetrics & Gynaecology*, 126(2), 138-141.

Wilkinson, J., Brison, D.R. Duffy, J.M.N., Farquhar, C.M., Lensen, S., Mastenbroek, S.,van Wely, M., & Vail, A., (2019b). Don't abandon RCTs in IVF. We don't even understand them, *Human Reproduction*, 34(11), 2093–2098.

Will, C., & Moreira, T., (2010). *Medical Proofs, Social Experiments: Clinical Trials in Shifting Contexts*. Ashgate, Farnham (UK) and Burlington (US).

Figure 1. Interviews with professionals

Professionals currently working in NHS fertility clinics	Other stakeholders (privately employed or from external organisations)
9 Senior embryologists	2 Embryology lab directors
9 Embryologists	2 Embryologists
4 Nurses	2 Medical directors
1 Medical director	2 Fertility consultants
2 Fertility consultants	4 Biotechnology company representatives
	2 Professional body representatives
	1 Bioethicist
	1 Patient advocate
	2 Clinical research professionals