

Time as an outcome measure in fertility-related clinical studies: long-awaited

Sesh K. Sunkara^{1,*}, Wenjing Zheng², Thomas D'Hooghe^{2,3,4}, Salvatore Longobardi⁵, and Jacky Boivin⁶

¹Division of Women's Health, Faculty of Life Sciences and Medicine, King's College London, London, UK ²Global Medical Affairs Fertility, Research and Development, Merck KGaA, Darmstadt, Germany ³Research Group Reproductive Medicine, Department of Development and Regeneration, Organ Systems, Group Biomedical Sciences, KU Leuven (University of Leuven), Leuven, Belgium ⁴Department of Obstetrics and Gynecology, Yale University, New Haven, CT, USA ⁵Global Clinical Development, Merck Serono S.p.A, Rome, Italy, an affiliate of Merck KGaA, Darmstadt, Germany ⁶School of Psychology, College of Biomedical and Life Sciences, Cardiff University, Cardiff, UK

*Correspondence address. Division of Women's Health, Faculty of Life Sciences and Medicine, King's College London, London WC2R 2LS, UK. E-mail: sesh.sunkara@kcl.ac.uk

Submitted on October 30, 2019; resubmitted on April 27, 2020; editorial decision on May 4, 2020

ABSTRACT: Time taken to achieve a live birth is an important consideration that is central to managing patient expectations during infertility treatment. However, time-related endpoints are not reported as standard in the majority of fertility-related clinical studies and there is no internationally recognized consensus definition for such endpoints. There is, therefore, a need for meaningful discussions around the selection of appropriate time-related treatment outcome measures for studies evaluating fertility treatments that will be relevant to diverse stakeholders (e.g. patients, healthcare professionals, clinical scientists, authorities and industry). Here, we provide a proposal for the evaluation of time-related outcome measures in fertility-related clinical studies, alongside associated definitions.

Key words: clinical trial design / time / medically assisted reproduction / time to live birth / statistical study design / outcomes / study design / trial design / endpoints

Introduction

Time taken to achieve a live birth is an important consideration when selecting fertility treatment, in particular because both maternal and paternal age at the point of attempting childbearing are increasing (ACOG, 2014; Steiner and Jukic, 2016; Hurley and DeFranco, 2017; Bergh *et al.*, 2019). Consequently, patients are older when they realize a need for fertility treatment (Baird *et al.*, 2005; Gleicher *et al.*, 2007; Lesthaeghe, 2010; Harris *et al.*, 2011; Mills *et al.*, 2011) and the biological timeframe to achieve a live birth is shortening.

As knowledge of medically assisted reproduction (MAR) has expanded, the scope to individualize fertility treatment to optimize outcomes based on patient characteristics has increased, enabling more patients to achieve their goal of having a child. However, once patients enter treatment there is no guarantee of immediate success, with the likelihood of pregnancy and subsequent live birth within a specific time period dependent on patient characteristics and treatment decisions (Smith *et al.*, 2015; Chambers *et al.*, 2017; O'Brien *et al.*, 2017). The likelihood of achieving a live birth increases cumulatively with each

additional attempt, and the majority of patients may now achieve treatment success—but only if they have sufficient time for multiple treatment cycles before natural fecundity declines markedly. In addition, there may be age cut-offs for provision of state-funded fertility treatments. For example, the National Institute for Health and Care Excellence (NICE) fertility guideline in the UK has a cut-off of 42 years for state-funded IVF treatment and delays (including waiting lists) may result in women not being eligible for treatment or for a full course of treatment (Dolan and Rudisill, 2015). Longer treatment duration before achieving a live birth will also increase costs and reduce cost-effectiveness both directly and indirectly, increasing the financial burden on patients (if they are self-funding their treatment) and/or providers (Eijkemans *et al.*, 2017). Prolonged time in fertility treatment is stressful (Wu *et al.*, 2013), and any financial burden placed on patients may further increase the stress.

These concerns, although not exhaustive, emphasize the need to achieve a live birth in as timely a manner as possible. However, to date, time has not been prioritized as a construct of interest in studies of fertility treatment. Instead, the majority of commonly used

treatment measures are based on the achievement (or not) of a fertility event of interest (e.g. pregnancy, live birth, etc.).

Here, we first discuss the evidence from the current literature on each aspect of a time-related treatment measure, based on a systematic search. Then we provide suggestions on what might be the optimal choice for each aspect and propose a time-related treatment measure for use in future fertility clinical research.

Literature review of time-related measures in fertility studies

A systematic search of the literature was performed to understand how time-related treatment outcomes were reported in clinical fertility studies (Table I). In the studies identified (14 randomized controlled trials [RCTs], 19 retrospective cohort studies and 9 prospective cohort studies), no single time-related outcome measure was consistently used, although the most commonly cited measure was 'time to pregnancy'. According to International Committee Monitoring Assisted Reproductive Technologies [ICMART], 'time to pregnancy' is defined as 'the time taken to establish a pregnancy, measured in months or in numbers of menstrual cycles' (Zegers-Hochschild *et al.*, 2017). However, this definition does not include the type of pregnancy to be evaluated (biochemical, clinical or ongoing), or any details of the start time point (STP) for measurement of the obstetric outcome or the end time point (ETP) (i.e. the time at which pregnancy should be confirmed). There is also no guidance on how this outcome should be evaluated (i.e. how to analyse it statistically or report it). 'Time to pregnancy leading to live birth' will be included in the core outcome set for fertility trials developed during the Core Outcome Measures for Infertility Trials project (Duffy *et al.*, 2018). The COMMIT project

was initiated because many different outcomes and outcome measures were reported in RCTs evaluating interventions for infertility, and it was proposed that this variability made it much more difficult to compare and combine individual trials to inform clinical practice.

There is, therefore, no well-defined or justified treatment time measure identified that can be put forward for use in future clinical research. In addition, the heterogeneity and lack of clear definitions mean that comparison between the studies or pooling of time-related treatment measures for meta-analysis might not be possible. This heterogeneity in outcomes has previously been identified as an issue for fertility studies, with a previous review identifying many different combinations of numerator and denominator in RCTs evaluating IVF treatment, with no single treatment measure used by a majority of studies (Wilkinson *et al.*, 2016). In addition, only around one-quarter of studies we identified used 'live birth' in their time-related treatment measure despite expert consensus specifying live birth as the most important measure of fertility treatment success (Martins *et al.*, 2018). These issues highlight the need for education on suitable outcomes for use in clinical studies, as well as on their analysis.

Time-related treatment measures for fertility: time to consider

The ability to compare the effectiveness of treatments with respect to a time-related measure would enable treatment selection with the aim of reducing time to achieve a live birth. However, the comparison of the effectiveness of different treatments is only possible if there is a standard, widely accepted and clinically meaningful treatment measure related to achieving a successful obstetric outcome (Maheshwari *et al.*, 2015; Mol *et al.*, 2018). The availability of such an outcome measure would also

Table I Summary of systematic literature search criteria.

Database(s) searched	PubMed, MEDLINE, Embase
Cut-off date for publications	16 August 2019
Key words/search terms	ARTs, fertility treatment, ovarian stimulation, ovulation induction, IVF, intracytoplasmic sperm injections, infertility, subfertility/subfecundity/sterility, intrauterine insemination or intracytoplasmic sperm injection AND Time to or duration AND Birth/arrival/baby/pregnancy/delivery/conception
Excluded terms	HIV, cancer, neoplasm, timelapse
Screening criteria for inclusion^a	A clinical study in the context of infertility (population is infertile or had clinical risk factors for infertility, e.g. fibroids, polycystic ovary syndrome, etc.) and/or a population is undergoing fertility management or treatment (also including fertility referral and fertility investigation), with a treatment measure that contained duration of time as an outcome (e.g. time to pregnancy or time to live birth).
Screening criteria for exclusion^a	Non-clinical studies, including epidemiological incidence and prevalence studies, reviews, meta-analyses, case reports and case series.

Results were automatically filtered to include studies in humans and publications in English, and duplicates were removed.

^aBased on titles and abstracts.

better enable meta-analyses across studies by reducing the heterogeneity in pooled results, and thereby increasing the power to detect differences between treatments. This has already been done for a number of outcomes (e.g. the ICMART definition of cumulative live birth rate), and is particularly important in this context because a time-related measure is unlikely to be the primary endpoint evaluated in the majority of studies. Such studies are, therefore, unlikely to be statistically powered to detect differences in this endpoint, with meaningful treatment differences only observed through pooled analysis. Moreover, time to event analysis in meta-analyses are almost exclusively underpowered at present (Stocking et al., 2019), so more powerful analyses are imperative.

Definitions of time to pregnancy and live birth

Defining the STP

There is only one STP that is appropriate for prospective clinical trials, which is the time of randomization (Fig. 1). This is because randomization is the point in all treatment groups from which variation in outcomes could be produced. For non-randomized clinical studies,

including observational studies, the aim should be to mimic RCTs as closely as possible. This means that a uniform STP defined at a particular point in the treatment pathway for all non-randomized studies would imply a common study design (with respect to start of intervention/treatment) for all, regardless of the research question. The STP should, therefore, be either the time of randomization or, evoking the idea of a target trial, the time at which you would have randomized patients if it had been an RCT. This should coincide with the start of intervention/treatment in the study.

Across current literature, the STP for treatment measures is fairly homogeneous, with randomization being the start point for the majority of RCTs and, where stated, a clinical procedure or the start of MAR treatment for non-RCTs. In some instances, studies evaluate a time-related treatment measure with an STP that relies upon patient recall of, for example, 'first intercourse after delivery' and 'decision of desire to have a child' (Boujenah et al., 2016; Boltz et al., 2017; Cayan et al., 2017).

Defining the ETP

The ETP for any study should be selected based on the aims of the study, the treatment being evaluated and the outcome that is expected to be affected. As the desired outcome of fertility treatment is live birth, this should be considered as a preferred ETP. However,

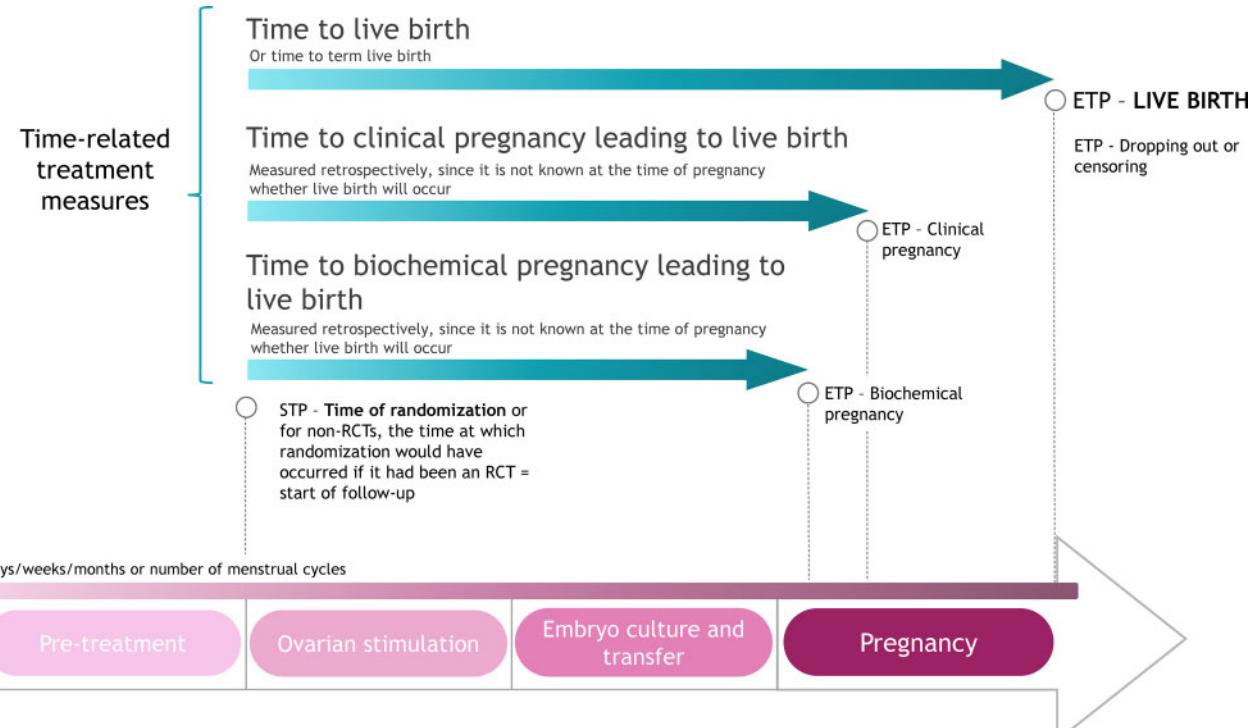


Figure 1. Summary of the proposed time-related treatment outcome measures for studies evaluating fertility treatment.

Biochemical pregnancy is defined as a pregnancy diagnosed only by the detection of beta hCG in serum or urine; clinical pregnancy is defined as a pregnancy diagnosed by ultrasonographic visualization of one or more gestational sacs or definitive clinical signs of pregnancy; live birth is defined as the complete expulsion or extraction from a woman of a product of fertilization, after 22 completed weeks of gestational age which, after such separation, breathes or shows any other evidence of life, such as heart beat, umbilical cord pulsation or definite movement of voluntary muscles, irrespective of whether the umbilical cord has been cut or the placenta is attached (Zegers-Hochschild et al., 2017). ETP, end time point; RCT, randomized controlled trial; STP, start time point.

the question becomes how to measure time to live birth in a way that does not erroneously favour preterm births, which would result in shorter times. Possibilities include measuring 'time to term live birth', or 'time to clinical pregnancy leading to live birth' (Fig. 1). It should be noted that it can only be determined whether or not a patient has had a positive outcome with respect to this latter measure in retrospect, since it is not known at the time of pregnancy (i.e. when there is a positive pregnancy test), whether or not a live birth will occur. Ongoing pregnancy rate is correlated with live birth rate and has been suggested as a surrogate measure when live birth is not recorded (Mol *et al.*, 2018); however, ongoing pregnancy as an outcome does not take into account later pregnancy losses and this may affect evaluation of time to treatment success. Furthermore, ongoing pregnancy is not defined in the latest ICMART Glossary (Zegers-Hochschild *et al.*, 2017), and is defined differently across studies, mostly corresponding to a pregnancy marked by gestational sac with positive foetal heart activity at 12th week of gestation (Kasius *et al.*, 2014; Dieamant *et al.*, 2017; Lensen *et al.*, 2018; Wang *et al.*, 2019).

There may be difficulties in reliably collecting pregnancy data at fixed times during a pregnancy outside of prospective studies, with most clinical pregnancies detected during a standard early-pregnancy dating scan. The inclusion of additional screening visits to a study will increase costs and the burden of the study on the patient and the healthcare provider. Where this is the case, we propose that researchers should opt for the most suitable ETP for their study. This option takes into account the fact that many investigators will be interested in evaluating the effect of an intervention on early (upstream) reproductive outcomes more closely related in time to the intervention under evaluation. Any later or downstream effect of a specific intervention on foetal loss, after the establishment of pregnancy with a positive foetal heartbeat, may be influenced predominantly by other foetal or maternal factors than the intervention itself, may be difficult to predict, hard to analyse in prospective research and impossible to properly evaluate, as the number of patients required would be too large to conduct an RCT. In this context, the assessment of biochemical pregnancy leading to live birth or clinical pregnancy with a positive foetal heartbeat leading to live birth may be attractive options, depending on when pregnancy data are collected during routine clinical practice. It is, however, crucial that the measurement times do not systematically differ between the arms of the study, as could occur during observational studies if ultrasound appointments were booked earlier for one treatment arm than for another. Similarly, the outcome could be biased if an unblinded outcome assessor measured the outcome in a more punctual fashion for their preferred treatment. Currently, available approaches in real-world data research may enable the evaluation of time to achieve these proposed outcomes in large numbers of patients. These assessments can be corrected (i.e. propensity scoring) for relevant baseline and treatment-related confounding factors, and early-pregnancy outcome data can be translated into (cumulative) live birth data.

Current literature displays heterogeneity with respect to the ETPs reported. ETPs of time-related treatment measures include 'time to conception/leading to live birth/ongoing pregnancy', 'time to (ongoing) pregnancy/leading to live birth' and 'time to live birth'. In many cases, the definition of the reproductive ETP (e.g. pregnancy or conception) is not provided. It also should be noted that the term 'conception' and its derivatives were removed from the most recent ICMART

glossary, owing to consensus to use scientifically recognized definitions such as 'fertilization', 'implantation', 'pregnancy' and 'live birth'.

Evaluating the time-related treatment measure

Although the aim of fertility treatment is to achieve a live birth, not all patients will accomplish this goal. A common mistake is to calculate time-related measures only in the subset of participants having a positive outcome. This induces selection bias, since it omits every participant for whom treatment did not work. A particularly poor treatment could appear to be very promising as a result of this error. It is, therefore, crucial that measurements include all randomized participants, regardless of whether they have a live birth or not. The follow-up times of all participants should be included until they either have the event or are censored. Censoring occurs when the event of interest is not observed because the follow-up period of the study has ended, the participant is lost to follow-up or else withdraws from the study. To this end, we would suggest presenting a Kaplan–Meier plot or cumulative incidence plot for the likelihood of achieving a clinical pregnancy with a foetal heartbeat that results in a safe live birth over time.

Time to event may be measured in number of days (i.e. weeks or months) or in number of menstrual or treatment cycles. Measuring time to event in number of days provides an estimate of the time taken to achieve the event of interest during treatment, taking into account multiple treatment cycles, the average wait times between those cycles and any other delays that might occur during the treatment (e.g. patient decision-making). An alternative approach is to measure time to event more discretely, for example, in number of menstrual/treatment cycles, which would account for any differences in the length of the treatment/menstrual cycles for individual patients and would be more applicable during counselling (see the section on patient counselling below). However, treatment cycles could vary in length based on the stimulation strategies (e.g. long GnRH agonist versus the short GnRH antagonist regimen) and measuring the time to event in cycles would not take account of delays before and/or during the treatment. Therefore, ideally, both approaches are used in combination to measure the time to event.

While it is reasonable to report numerical summaries of the time to event in each arm of a study (e.g. the time taken for some proportion of the group to achieve pregnancy leading to live birth) caution is needed before applying these summaries directly to new patients. This is because trials are generally not intended, and are thus generally not designed, for this purpose; their inclusion criteria are not specified so as to produce a representative sample of patients. However, relative effectiveness of the compared treatments can be evaluated in a trial using a ratio measure (Rothman *et al.*, 2008). This is because relative/ratio measures are generally reasonably stable when extrapolating from the selected, non-representative study populations to the patient population of interest, whereas absolute measures are not.

Hazard ratios (HR) with 95% CIs are the standard approach for summarizing time-related treatment effects. However, analyses based on HR require some assumptions. They assume that dropping out or censoring is unrelated to patient prognosis (Daya, 2005) and this is unlikely to be true for fertility treatment as patients may be more or less likely to drop out on the basis of advice regarding their chances of a successful outcome. Furthermore, it is known that when the treatment

is stressful or physically burdensome (e.g. there is a high risk of ovarian hyperstimulation syndrome [OHSS] owing to stimulation) the chance of treatment postponement increases (Gameiro *et al.*, 2012). This means that postponement cannot be disconnected from treatment as though it was a completely random factor (i.e. censoring in studies of fertility treatment is unlikely to be random but rather co-influenced by factors like risk of OHSS, poor ovarian response, and physical and/or psychological stress) (Daya, 2005). To make meeting this assumption more plausible, researchers should adjust for prognostic covariates in their analysis, for example, by using Cox regression. In addition, when calculating HR, the proportional hazards assumption is usually made. This assumption states that the relative difference between two treatment arms remains constant over time and if it does not hold true, then a summary based on a single HR will be misleading (Hernán, 2010).

Another approach (that may also be helpful for counselling) is to provide the data using a summary resembling median survival time (MST), a statistic commonly reported in oncology. MST is most commonly known as the time at which 50% of the cohort is expected to achieve the desired outcome (i.e. to become pregnant and go on to have live birth). For fertility studies, it may not be possible to calculate the time at which 50% become pregnant, as in many studies fewer than half the sample have live births. It might, therefore, be more appropriate to report survival rates at set time points, or even to calculate the time for 25% of the patients to become pregnant (i.e. time to quartile pregnancy [TQP]). However, it is important for clinical studies to show that outcomes might take more than one treatment cycle.

In current literature, methods commonly reported to calculate the data for the time-related endpoint include average/mean/median, but reported only in patients who achieve the outcome of interest (e.g. conception, pregnancy or birth); as well as Kaplan–Meier analysis, log-rank analysis and Cox proportional hazards analysis evaluated in the total analysis population. In most cases, duration of follow-up for analysis is set as a unit of time (such as months or years), although the number of treatment cycles is another approach (Diamond *et al.*, 2015). The use of agreed upon definitions for evaluating interventions that occur either outside of or during MAR treatment should enable comparison of interventions, enabling better shared decision-making to reduce the time to achieving a live birth. The question of how to make individualized treatment recommendations, for example, by combining the effect estimate from an RCT with a prediction modelling approach, is the subject of ongoing research (Kent *et al.*, 2018).

Use of time-related measures in counselling

While time-related measures are of value to the design of research in this field, it is ultimately patients who will be an important source of demand for time-related information when balancing considerations in treatment selection and decision-making. Time to event estimates could serve as a reference for patients matching the eligibility criteria of the studies. They may also be utilized in patient counselling with regards to setting expectations and facilitating shared treatment decision-making, taking into account the duration of treatment as one of the aspects to be considered.

However, conveying clinical data to patients brings inherent challenges. Language that the patient will understand should be used when communicating outcome data, including discussion of the applicability of the data to clinical practice. For example, the meaning and relation to time of HR may be difficult to convey. As an alternative, a summary value, such as MST, could be useful, but may also be misleading and so needs to be renamed for use in fertility treatment counselling. Care should also be taken in the presentation of other measures, such as TQP, which might encourage a single cycle view of fertility treatment.

The most valuable information to the patient regarding the length of the treatment should account for the ultimate goal of MAR—live birth. Multiple studies have demonstrated improvement in early or midway reproductive outcomes (i.e. positive biochemical pregnancy test, clinical pregnancy, ongoing pregnancy) with one intervention compared with the other; however, these failed to demonstrate the improvement in live birth. To this end, 'time to (term) live birth' or 'time to pregnancy leading to live birth' is encouraged as the preferred time-related treatment outcome.

When counselling patients on 'time to pregnancy leading to live birth', our proposals need to consider common decision-making heuristics that may bias understanding of probability/risk information (e.g. framing and anchoring effects, poor affective forecasting) (Elwyn *et al.*, 2011). It is also known that patients systematically over-evaluate their chances of pregnancy relative to what doctors predict, even when personal characteristics would predict poorer prognosis (e.g. age, years infertile) (Kowalcuk *et al.*, 2003). Therefore, future research is needed to evaluate how best to present such information to patients in a way to minimize bias in clinics but also to ensure knowledgeable and informed patient choice.

Conclusion

Currently, there is no standard definition of a time-related treatment measure for use in clinical studies evaluating fertility treatment and there is no consistency in the evaluation of these endpoints. We have, therefore, proposed time-related treatment measures (Fig. 2), which should be a call-to-action to ensure appropriate statistical analyses are used when evaluating time-related endpoints in clinical trials. Critically, our proposals for this time-related measure should be refined based on further discussions, and robustly validated by methodologists in regards to design and analysis, as these can be particularly complex and challenging. Overall, the topic of time-related treatment measures for fertility-related clinical studies is important for healthcare professionals and patients, and further discussion is warranted to validate appropriate measures for future use.

Acknowledgements

Medical writing assistance was provided by Alexander Jones, inScience Communications, London, UK, and funded by Merck KGaA, Darmstadt, Germany. We would also like to thank Alexander Jones (AJ) for assistance with the design and conduct of the systematic evidence search, data extraction and reference screening.

- For prospective randomized clinical studies, time of randomization should be the start point for the time-related treatment measure.
- For non-randomized clinical studies, the start point should be time at which patients would have been randomized in a randomized controlled trial.
- The end time point should be selected according to the aims of the study, the treatment being evaluated, and the expected (or a closely related) outcome of the intervention.
- Live birth should be the preferred end time point; however, ongoing pregnancy can also be considered, as this is strongly correlated with live birth.
- “Biochemical pregnancy leading to live birth” or “clinical pregnancy with a positive foetal heartbeat leading to live birth” are suggested endpoints when clinical pregnancy is detected during a standard early pregnancy scan.
- Time measures should not differ between the arms of the study and should include all randomized patients, even if they do not achieve the outcome of interest.
- A Kaplan–Meier or cumulative incidence plot could show the likelihood over time of achieving clinical pregnancy with a foetal heartbeat leading to live birth.
- Cox regression analysis is encouraged, to adjust for prognostic covariates and when drop-out or censoring is directly related to patient prognosis. Cox regression should be changed to other analyses (e.g., logistic analysis) if the assumptions are seriously violated.
- Time-related outcome measures may be beneficial for counselling patients about treatment choices that have comparable reproductive outcomes but different times to pregnancy leading to live birth.

Figure 2. Key points on the proposed time-related treatment measures for use in future fertility clinical research.

Authors' roles

S.K.S. was involved in design of the search and reference screening, and drafted, critically reviewed and approved the manuscript. J.B. and T.D.H. were involved in the study concept and design, and drafted, critically reviewed and approved the manuscript. S.L. and W.Z. critically reviewed and approved the manuscript.

Funding

Merck KGaA (Darmstadt, Germany) funded the preparation of this article, including independent medical writing assistance and literature review.

Conflict of interest

W.Z. and T.D.H. are employees of Merck KGaA, Darmstadt, Germany. S.L. is employee of Merck Serono S.p.A, Rome, Italy, an affiliate of Merck KGaA, Darmstadt, Germany. J.B. reports speaker or consultancy fees from Merck KGaA, Merck AB, Theramex, Ferring Pharmaceuticals A/S and grants from Merck Serono Ltd outside the submitted work. S.K.S. has lectured in symposia by Merck, MSD, Ferring and Pharmasure.

References

ACOG. Female age-related fertility decline. Committee Opinion No. 589. *Fertil Steril* 2014; **101**:633–634.

Baird DT, Collins J, Egozcue J, Evers LH, Gianaroli L, Leridon H, Sunde A, Templeton A, Van Steirteghem A, Cohen J et al. Fertility and ageing. *Hum Reprod Update* 2005; **11**:261–276.

Bergh C, Pinborg A, Wennerholm UB. Parental age and child outcomes. *Fertil Steril* 2019; **111**:1036–1046.

Boltz MW, Sanders JN, Simonsen SE, Stanford JB. Fertility treatment, use of in vitro fertilization, and time to live birth based on initial provider type. *J Am Board Fam Med* 2017; **30**:230–238.

Boujenah J, Hugues JN, Sifer C, Cedrin-Durnerin I, Bricou A, Poncelet C. Second live birth after undergoing assisted reproductive technology in women operated on for endometriosis. *Fertil Steril* 2016; **105**:129–133.

Cayan S, Sahin S, Akbay E. Paternity rates and time to conception in adolescents with varicocele undergoing microsurgical varicocele repair vs observation only: a single institution experience with 408 patients. *J Urol* 2017; **198**:195–201.

Chambers GM, Paul RC, Harris K, Fitzgerald O, Boothroyd CV, Rombauts L, Chapman MG, Jorm L. Assisted reproductive technology in Australia and New Zealand: cumulative live birth rates as measures of success. *Med J Aust* 2017; **207**:114–118.

Daya S. Life table (survival) analysis to generate cumulative pregnancy rates in assisted reproduction: are we overestimating our success rates? *Hum Reprod* 2005; **20**:1135–1143.

Diamond MP, Legro RS, Coutifaris C, Alvero R, Robinson RD, Casson P, Christman GM, Ager J, Huang H, Hansen KR et al. Letrozole, gonadotropin, or clomiphene for unexplained infertility. *N Engl J Med* 2015; **373**:1230–1240.

Dieamant F, Petersen CG, Mauri AL, Comar V, Mattila M, Vagnini LD, Renzi A, Petersen B, Ricci J, Oliveira JBA et al. Single versus sequential culture medium: which is better at improving ongoing pregnancy rates? A systematic review and meta-analysis. *JBRA Assist Reprod* 2017; **21**:240–246.

Dolan P, Rudisill C. Babies in waiting: why increasing the IVF age cut-off might lead to fewer wanted pregnancies in the presence of procrastination. *Health Policy* 2015; **119**:174–179.

Duffy JMN, Legro RS, Farquhar CM, Lensen S, Mol BW, Niederberger C, Ng EHY, van Wely M, Repping S, Strandell A et al.; COMMIT: Core Outcomes Measures for Infertility Trials. A protocol developing, disseminating and implementing a core outcome set for infertility. *Hum Reprod Open* 2018; **2018**:hoj007.

Eijkemans MJC, Kersten FAM, Lintsen AME, Hunault CC, Bouwmans CAM, Roijen LH, Habbema JDF, Braat DDM. Cost-effectiveness of 'immediate IVF' versus 'delayed IVF': a prospective study. *Hum Reprod* 2017; **32**:999–1008.

Elwyn G, Stiel M, Durand MA, Boivin J. The design of patient decision support interventions: addressing the theory-practice gap. *J Eval Clin Pract* 2011; **17**:565–574.

Gameiro S, Boivin J, Peronace L, Verhaak CM. Why do patients discontinue fertility treatment? A systematic review of reasons and predictors of discontinuation in fertility treatment. *Hum Reprod Update* 2012; **18**:652–669.

Gleicher N, Weghofer A, Barad D. Too old for IVF: are we discriminating against older women? *J Assist Reprod Genet* 2007; **24**:639–644.

Harris ID, Fronczak C, Roth L, Meacham RB. Fertility and the aging male. *Rev Urol* 2011; **13**:e184–e190.

Hernán MA. The hazards of hazard ratios. *Epidemiology* 2010; **21**:13–15.

Hurley EG, DeFranco EA. Influence of paternal age on perinatal outcomes. *Am J Obstet Gynecol* 2017; **217**:566.e1–566.e6.

Kasius A, Smit JG, Torrance HL, Eijkemans MJ, Mol BW, Opmeer BC, Broekmans FJ. Endometrial thickness and pregnancy rates after IVF: a systematic review and meta-analysis. *Hum Reprod Update* 2014; **20**:530–541.

Kent DM, Steyerberg E, van Klaveren D. Personalized evidence based medicine: predictive approaches to heterogeneous treatment effects. *BMJ* 2018; **363**:k4245.

Kowalcik I, Kasimzade T, Huber G. Expectations for success in fertility treatment involving assisted reproduction. *Arch Gynecol Obstet* 2003; **268**:78–81.

Lensen SF, Wilkinson J, Leijdekkers JA, La Marca A, Mol BWJ, Marjoribanks J, Torrance H, Broekmans FJ. Individualised gonadotropin dose selection using markers of ovarian reserve for women undergoing in vitro fertilisation plus intracytoplasmic sperm injection (IVF/ICSI). *Cochrane Database Syst Rev* 2018; **2**:CD012693.

Lesthaeghe R. The unfolding story of the second demographic transition. *Popul Dev Rev* 2010; **36**:211–251.

Maheshwari A, McLernon D, Bhattacharya S. Cumulative live birth rate: time for a consensus? *Hum Reprod* 2015; **30**:2703–2707.

Martins WP, Niederberger C, Nastri CO, Racowsky C. Making evidence-based decisions in reproductive medicine. *Fertil Steril* 2018; **110**:1227–1230.

Mills M, Rindfuss RR, McDonald P, te Velde E. Why do people postpone parenthood? Reasons and social policy incentives. *Hum Reprod Update* 2011; **17**:848–860.

Mol BW, Bossuyt PM, Sunkara SK, Garcia Velasco JA, Venetis C, Sakkas D, Lundin K, Simon C, Taylor HS, Wan R et al. Personalized ovarian stimulation for assisted reproductive technology: study design considerations to move from hype to added value for patients. *Fertil Steril* 2018; **109**:968–979.

O'Brien YM, Ryan M, Martyn F, Wingfield MB. A retrospective study of the effect of increasing age on success rates of assisted reproductive technology. *Int J Gynaecol Obstet* 2017; **138**:42–46.

Rothman KJ, Lash TL, Greenland S. *Modern Epidemiology*. Philadelphia, PA: Lippincott Williams & Wilkins, 2008.

Smith A, Tilling K, Nelson SM, Lawlor DA. Live-Birth rate associated with repeat in vitro fertilization treatment cycles. *JAMA* 2015; **314**: 2654–2662.

Steiner AZ, Jukic AM. Impact of female age and nulligravidity on fecundity in an older reproductive age cohort. *Fertil Steril* 2016; **105**: 1584–1588.e1.

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. *Hum Reprod* 2019; **34**: 659–665.

Wang R, Danhof NA, Tjon-Kon-Fat RI, Eijkemans MJ, Bossuyt PM, Mochtar MH, van der Veen F, Bhattacharya S, Mol BWJ, van Wely M. Interventions for unexplained infertility: a systematic review and network meta-analysis. *Cochrane Database Syst Rev* 2019; **9**: CD012692.

Wilkinson J, Roberts SA, Showell M, Brison DR, Vail A. No common denominator: a review of outcome measures in IVF RCTs. *Hum Reprod* 2016; **31**:2714–2722.

Wu AK, Elliott P, Katz PP, Smith JF. Time costs of fertility care: the hidden hardship of building a family. *Fertil Steril* 2013; **99**: 2025–2030.

Zegers-Hochschild F, Adamson GD, Dyer S, Racowsky C, de Mouzon J, Sokol R, Rienzi L, Sunde A, Schmidt L, Cooke ID et al. The international glossary on infertility and fertility care, 2017. *Fertil Steril* 2017; **108**:393–406.