Statistical Analysis Plan

eValuating iDA Selection Ability. The VISA study.

Will embryo selection through use of artificial intelligence (iDA) perform equally compared to blastocyst scoring?

2022-11-06

Author

Name/Title:
Nils-Gunnar Pehrsson / Senior Statistician, CEO, Statistiska Konsultgruppen

Nils-G. Pehrsson 8 Nov 2022

Signature Date

Approvals

Name/Title:
Peter Illingworth / Principal Investigator, Virtus Health

Signature Date 9/11/22

Name/Title:
Thorir Hardarson / Vitrolife AB

Signature Date

Revisions

Version	Description of Changes	Date
1	A general overview and adjustments according to the latest CIP version (Sweden)	2022-08-17
2	10 clinics in 2 countries have been added	2022-08-17
3	Changes in which baseline parameters should be used	2022-08-17
4		

Table of Contents

LIST OF ABBREVIATIONS

Abbreviation	Definition
FHB	Fetal heartbeat
iDA	iDAScore® which is the commercial name for the deep learning tool investigated
hCG	Human chorionic gonadotropin
IU	International units
Hpi	Hours post insemination
CIP	Clinical investigation protocol
RCT	Randomised controlled trial
CI	Confidence interval
ITT	Intention-to-treat
FAS	Full analysis dataset
PP	Per-protocol
BMI	Body mass index
FSH	Follicle stimulation hormone
LH	Luteinizing hormone
PN	Pronuclei
GnRH	Gonadotropin-releasing hormone
ICSI	Intracytoplasmic sperm injection
IVF	In-vitro fertilisation
AI	Artificial intelligence
AE	Adverse event
OR	Odds ratio
RR	Relative risk

SAE	Serious adverse event
DSMB	Data safety and monitoring board

1 STUDY DETAILS

Study Objectives

The primary objective for this study is to investigate whether selection of a single blastocyst for transfer using the deep learning-based support tool called iDAScore results in a non-inferior clinical pregnancy rate compared to when the selection is performed by trained embryologists using conventional morphology only. Clinical pregnancy is defined in this study as the detection of a fetal heartbeat (FHB) by ultrasound after 42 days of gestation (week 6) (including ectopic pregnancies, as per the ICMART definition).

The secondary objectives for this study are:

To investigate whether blastocyst selection supported by iDAScore results in non-inferior rates of the below parameters compared to when embryo selection performed by trained embryologists using conventional morphology only:

1. Live birth rate* (defined as the number of patients with at least one live birth after 22 completed weeks of gestation. If the exact gestational age is not known then a birth weight of ≥500gr can be used as a cut-off.

2. Positive β-hCG rate (defined as the number of patients with a positive β-hCG, determined by a hCG measurement from a blood sample or using urinary sticks).

3. Rate of non-viable pregnancies (defined as the difference between number of clinical pregnancies (excluding ectopic pregnancies) and number of positive β-hCG pregnancies)

4. Ongoing pregnancy rate* (defined as the number of patients with a viable pregnancy at ≥12 weeks of gestation)

*These secondary parameters will not be included in the first statistical analysis and report as these will not be available until approximately 12 months after the initial data collection.

Study Design

This is a non-inferiority, prospective parallel group, double blind, randomized controlled trial in 15 IVF centres in 4 countries (Australia, Denmark, Sweden and United Kingdom).

Treatment Groups

Patients will be randomly allocated (1:1) to two different study groups if at least two embryos have developed to the blastocysts stage (Gardner score 2 or higher):

A. **Control group: Embryo selection by standard morphologic criteria.**

- The embryo for transfer will be selected by the embryologist based on the morphologic appearances on day 5, according to the Gardner criteria (Gardner et al., 2000) using the ranking guideline (Appendix 1, in CIP).
- Regardless of whether a transfer takes place any embryos that fulfil the criteria applied in the laboratory for cryopreservation will be cryopreserved (Appendix 1 in the clinical investigation protocol (CIP)). If there is doubt about whether an embryo is suitable for cryopreservation, embryos may be held over to day 6 and the decision will be made then.
- The order of rewarming and transfer of cryopreserved embryos for later transfer will be made according to Appendix 1 in the CIP based on the all the information that is available by day 6.

B. **Treatment group: Embryo selection supported by iDAScore**.

- The time-lapse videos will be analysed by iDA at 114-118 hours post insemination (hpi) and the embryo with the highest iDA score will be selected for fresh transfer on day 5.
- Any remaining viable embryos in this group will be cryopreserved if they have either:

  - reached Gardner Grade 3 or beyond AND would normally be cryopreserved or;
  - have reached Gardner Grade 3 AND achieved a score on iDA of 5 or more.
- All other embryos will be cultured until day 6, re-scored at 138-142 hpi and according to the above criteria.
- The warming and transfer of embryos will be performed based on the ranking of the iDA score across the two days. The first embryo to be warmed will be the one with the highest iDAScore. If this embryo does not survive warming and is not suitable for transfer, the next embryo to be warmed will be selected based on the iDAScore, until an embryo is warmed and is suitable for transfer.

Embryo transfer will be performed using the clinic's routine methods. Luteal support will be administered using the standard protocol of each clinic. Both the treating clinician and the patient will remain blinded to the randomization outcome until after the first embryo transfer has been completed.

Sample Size calculation

It is estimated from the results in clinics that clinical pregnancy is estimated to be 35.4% for trained embryologists. If non-inferiority margin is defined as - 5%, the lower limit of the two-sided 95% confidence interval (CI) for the difference between iDAScore group and Trained embryologist group shall not be less than -5% with a probability of 90% (β=10%), with an estimation of 5% or more clinical pregnancies in iDAScore group, 494 women per randomization group is needed to show non- inferiority with two-sided Farrington-Manning test. For protection against a 5% loss to follow-up, 1040 patients in total, 520 per group, are needed for recruitment.

2 STUDY POPULATIONS

Definition of Study Populations

2.1.1 Intent-to-Treat Population

All randomized subjects will be included in the Intent-to-Treat (ITT) population. Subjects will be analysed according to randomized group.

2.1.2 Full Analysis Set

All randomized subjects with measurement of primary efficacy variable will be included in the Full Analysis Set (FAS). Subjects will be analysed according to randomized group.

2.1.3 Per-Protocol Population

All randomized subjects with no significant protocol violations will be included in the Per Protocol (PP) population. Subjects will be analysed according to actual embryo selection method. The Per Protocol population will be defined during the clean file meeting before the database lock without knowledge to which randomised group the patient belongs to.

2.1.4 Safety Population

All enrolled subjects who started one of the study embryo selections methods (i.e., in patients with a minimum of two early blastocysts) will be included in the safety population.

The final decisions regarding all the above study population will be taken at the Clean File meeting before the database lock.

3 POOLABILITY OF INVESTIGATIVE SITES

The data from all investigative sites will be pooled based on the assumption of clinical comparability: the sites used a common protocol; the sponsor adequately monitored the study to assure protocol compliance; and the data gathering and validation mechanisms were the same across all study sites.

Analyses to justify pooling will include the following:

- The primary endpoint will be presented by site: Mean difference in percentages of clinical pregnancy with 95% CI.

- The justification for pooling all the data to estimate a common effect across study sites requires the homogeneity of response across study sites. An exact Pearson Chi-square test of the proportions of clinical pregnancies over sites will be generated to test whether the investigational sites differ with respect to primary study on the ITT population (only the site will be included in this analysis model). The test of homogeneity of response will be based on a two-sided significance test at the 0.10 level of significance.

- If the sites differ by this test then a second analysis will be done including study site and all baseline characteristics that have a $p<0.10$ in the analysis of baseline variables by site to understand if the imbalance in the primary endpoint between sites is related to an imbalance in baseline characteristic.

The analysis of study sites may require the formation of pseudo-sites because the small study sites will not provide appropriate information to allow the analysis above. Study sites with fewer the 16 subjects will be

combined into pseudo-sites for the poolability analysis using the following method. The smallest study site with less than 16 subjects will be combined with the next smallest study site. These two sites may be combined with a third site if the combined number of patients remain less than 16.

4 STUDY VARIABLES

Baseline Variables

4.1.1 Demographics and Baseline Characteristics

1. Age (maternal*/paternal), continuous variables.
2. Reason for infertility (couple), categorical variable.
3. Height and weight (mat.), continuous variables
4. BMI (mat.), continuous variable
5. Type of menstruation (mat.), dichotomous variable
6. Number of previous stimulated IVF cycles leading to oocyte pick-up (couple)*, ordered categorical variable (7 categories)
7. Previous pregnancies in current relationship, ordered categorical variable (5 categories)

Treatment variables

1. FSH starting dosage, continuous variable
2. FSH total dosage, continuous variable
3. GnRH downregulation (agonist/ antagonist)
4. Source of sperm, categorical variable (6 categories)
5. Duration of ovarian stimulation, continuous variable.
6. Number of oocytes*, both as continuous variable and ordered categorical variable (7 categories)

7. Method of fertilization (ICSI/Standard IVF/Combined)* Categorical variable
8. Number of normally fertilized oocytes (2PN)*, continuous variable
9. Number of blastocysts at Gardner Scale 2 or beyond by day 5 (Early blastocysts), continuous variable
10. Proportion of cycles where the embryo selected by the embryologist had the highest iDAScore (Only in the iDAScore group).
11. Number of cryopreserved embryos on day 5 and 6, continuous variable
12. Morphological score of the transferred embryo, continuous variable
13. iDAScore® (treatment group) of the transferred embryo, continuous variable
14. Type (categorical variable) and duration (continuous variable) of luteal phase support

*) These variables are used in a deterministic minimization procedure together with center during randomization. All variables are entered as Categorical variables: Age (5 cat.), number of earlier IVF cycles (7 cat.), number of oocytes (7 categories), fertilization method (3 cat.), number of 2PN oocytes (7 categories) and center (14-15 centers).

Efficacy Variables

4.1.2 Primary Efficacy Variable

Primary efficacy variable will be Clinical pregnancy with fetal heartbeat after the first embryo transfer per randomised patient. If no fresh nor frozen embryo can be transferred, then for the ITT population a negative clinical pregnancy is given for these patients.

4.1.3 Secondary efficacy Variables

Secondary efficacy variables will be the following outcome variables:

- Live birth (Y/N). Will be analysed in a second database lock.
  With pass criteria non-inferiority analysis, non-inferiority margin -5%.

- Positive hCG rate per randomized patient
  With pass criteria non-inferiority analysis, non-inferiority margin -5%.
- Non-viable intrauterine pregnancies

Safety Variables

4.1.4 Adverse Events (AE)

The occurrence of adverse events is documented in the eCRF through a separate AE module.

5 STATISTICAL METHODOLOGY

General Statistical Methodology

Primary and all secondary analyses will be performed on ITT population. Complementary analyses will be performed on full analysis set (FAS) and on the per-protocol (PP) population. The primary statistical analyses will be calculation of the mean percentage difference with two-sided 95% confidence interval (CI) regarding the primary efficacy variable Clinical Pregnancy with fetal heartbeat after the first embryo transfer between the iDA group and the trained embryologist (standard) group unadjusted with Farrington-Manning 95% CI. If the lower limit of this 95% CI is larger than -5%, the non-inferiority margin, then non-inferiority is achieved. If non-inferiority is achieved, then we will test primary efficacy variable for superiority with two-sided Fisher's exact test.

Unadjusted and adjusted relative risk (RR) between the two groups with 95% CI will be calculated with Poisson regression model with a robust error variance, (see reference 2) for primary and secondary efficacy variables.

Regarding multiplicity for the primary analysis (clinical pregnancy) and the first two secondary analyses (live birth) and Positive hCG a fixed – sequence test will be applied. (see reference 3)

If non-inferiority is achieved in primary analysis then the probability mass from the primary analysis will be transferred to the first secondary analysis, live birth.

If non-inferiority is achieved also for live birth then the non-inferiority result for live birth will also be confirmative, and the probability mass will be transferred to the second secondary analysis of positive hCG rate. If non-inferiority is achieved also for positive hCG rate birth, then the non-inferiority results for positive hCG rate will also be confirmative

For the other secondary variables, the p-values will be given for descriptive purpose and no multiplicity adjustment will be performed. These analyses will be considered exploratory.

For unadjusted comparison between the two randomized groups the following tests will be performed: Fisher's exact test for dichotomous variables,

Fisher's non-parametric permutation test for the mean difference between two independent samples for continuous variables,

Mantel-Haenszel chi-square test for ordered categorical variables and

Pearson chi-square test for non-ordered categorical variables.

Dichotomous data will be expressed as numbers and percentages. Continuous variables will be described with mean, standard deviation, median, quartile 25%, quartile 75%, minimum and maximum.

If baseline confounders, variables that differ statistically and clinically between the randomized groups and known to predict primary outcome variable, are found then complementary analyses will be performed adjusted for these baseline variables.

All tests will be two-tailed and conducted at 0.05 significance level.

All safety analyses will be descriptive by performed embryo selection method.

All analyses will be performed by using SAS® v9.4 (Cary, NC).

Patient Disposition and Data Sets Analysed

The number of subjects included in each of the ITT, FAS, PP and safety populations will be summarized for each treatment group and overall. The number and percentage of subjects randomized and treated will be presented. Subjects who completed the study and subjects who withdrew from study prematurely will also be presented with a breakdown of the reasons for withdrawal by treatment group for the ITT, FAS, PP and safety populations.

Protocol Violations/Deviations

Major protocol deviations are those that are considered to influence the analysis. A list of protocol deviations that were anticipated before the start of the trial was generated. Once the data collection has been finalized and before database lock the steering committee will review all the possible deviations and if needed create additional categories.

The number of patients with major protocol deviations will be summarized per treatment group.

Demographics and Baseline Characteristics

Demographics and baseline characteristics will be summarized by treatment group for the ITT, FAS and PP populations and analysed according to the methods described in section “General Statistical Methodology” above.

Main Treatment variables

Main treatment variables will be summarized by treatment group for the FAS and PP populations and analysed according to the methods described in section “General Statistical Methodology” above.

Efficacy Analyses

Primary Efficacy Analysis

The primary statistical analyses will be calculation of the mean percentage difference with two-sided 95% confidence interval (CI) regarding the primary efficacy variable Clinical Pregnancy with fetal heartbeat after the first embryo transfer between the iDA group and the trained embryologist (standard) group unadjusted with Farrington-Manning 95% CI. If the lower limit of this 95% CI is larger than -5%, the non-inferiority margin, then non-inferiority is achieved. This analysis will be applied on the ITT population (primary) and PP population (sensitivity analysis). Any inconsistencies in these analyses will be discussed in the study report. Unadjusted relative risk (RR) with 95% CI will also be calculated. The Clinical pregnancy rate with exact 95% CI will be calculated for each of the randomised groups.

The following sensitivity analysis will also be performed adjusted for centre and stratification (minimization) variables:

To account for the lack of independence introduced by both the potential clustering within centers and the optimal randomization process (minimization), we will also estimate the difference in the clinical pregnancy probability between the two randomized groups with a 95% CI using a linear mixed effects regression model (with identity link function) with center as a random effect and fixed effects for the following variables used in the randomization:

- Woman’s age:

< 25; 25 <= Age < 30; 30 <= Age < 35; 35 <= Age < 40; 5. 40 <= Age

- Number of previous stimulated IVF cycles leading to oocyte pick-up

1-5; 6 – 10; 11 – 15; 16 – 20; 21 – 25; 26 -> ....

- Number of oocytes:
- *2 – 5; 6 – 10; 11 – 15; 16 – 20; 21 - 25; 26 -> ....*
- Fertilization method:

IVF for all; ICSI for all; ICSI for some

- Number of 2PN oocytes
- *2 – 5; 6 – 10; 11 – 15; 16 – 20; 21 - 25; 26 -> ....*
- Center.
  Adjustment with centre as random effect, fertilization method as class variable and the other four as continuous variables.

To account for potential non-constant variance of errors in this model, we will estimate the empirical covariance matrix using the heteroskedasticity consistent method (HC3) as proposed by MacKinnon and White (see reference 1).

Adjusted relative risk (RR) with 95% CI will also be calculated adjusted for the minimization variables using Poisson regression model with a robust error variance. .

If non-inferiority is achieved, then we will test the primary efficacy variable for superiority with a two-sided Fisher's exact test. The primary analysis will also be performed per center. If baseline confounders are found the adjusted analyses will also be performed adjusted for these variables.

5.1.1 Secondary Efficacy Analyses

The secondary efficacy analyses will be the analysis of the secondary efficacy variables defined in section 4.1.3 “Secondary efficacy variables” according to the methods given in section 5, under “General Statistical Methodology”.

5.1.2 Exploratory Efficacy analyses

- Compare the clinical pregnancy in the iDAScore groups that doesn’t agree with the embryologist with clinical pregnancy in the agreement group in the iDAScore arm, with the same analysis as the primary analysis.
- Calculate the percent agreement in the iDAScore arm between the iDAScore and the embryologist regarding clinical pregnancy.

***5.1.3** Pre-specified subgroup analyses*

Pre-specified subgroup analysis will be performed on the primary and important secondary variables between the two randomized groups for the following baseline subgroup:

- Women older than 35 years.

5.1.4 Exploratory interaction analyses

Exploratory interaction analyses between the two randomized group and the following baseline variables regarding analyses of primary and selected secondary variables:

- Maternal age
- Number of blastocysts at Gardner Scale 2 or beyond by day 5
- Freeze all

Model: Outcome = Randomized_Group Baseline_var Randomized_Group *Baseline_var

These analyses should be performed with no other adjustments with Poisson regression model with a robust error variance.

If interaction p-value <0.10 then subgroups analysis will follow regarding primary and selected secondary analyses in subgroups of the baseline variable.

Statistical Analysis Plan (SAP) for the sub study:

Comparing time used for embryo evaluation between the conventional morphology group (control) and iDAScore (treatment) group.

Fisher's non-parametric permutation test for the mean difference between two paired samples will be used to analyze time used for embryo evaluation between the two methods used on the same individuals. For each method mean, SD, Median, minimum and maximum will be given. For the difference between the two methods mean with 95% CI, SD, median, minimum and maximum will be given. Six subjects are the minimum number to be included from each centre.

Separate analyses will also be performed for each site and divided in three categories of number of embryos.

Follow-up for primary efficacy analysis

All patients undergoing a freeze-all cycle for any clinical indication, are expected to return for an embryo transfer within 3 months after their oocyte collection. All freeze-all patients that have not returned for their first rewarmed blastocyst 3 months after the last patient was randomized will be treated as a protocol violations.

Safety Analyses

5.1.5 Adverse Events

Only treatment-emergent AEs will be included in the summaries for safety population.

A summary of subjects reporting at least one of the following AEs will be presented in an overview table:

- Any AE
- Any SAE
- Any treatment-related AE
- Any treatment-related SAE
- Any AE leading to discontinuation
- *Any device related events*
- Death

Summaries per SOC and PT presenting n (%) of AEs and n (%) of subjects with at least one AE will be provided for:

- All AEs (includes all serious and non-serious AEs)
- All AEs by maximum reported intensity
- All AEs by causality
- All SAEs
- All AEs leading to discontinuation

6 DSMB ANALYSES AND INTERIM ANALYSES

Data Safety Monitoring Board analyses.

An independent DSMB has been appointed to follow the safety and efficacy monitoring as well as the overall conduct of the study. The board consists of a statistician and a medically knowledgeable person

unrelated to the study. The role of the DSMB has been set out in a separate charter and the members will hold regular meetings where the study efficacy and safety will be assessed and if necessary, suggestions for changes in study protocol. For early termination of efficacy for benefit (clinical pregnancy substantially higher in iDA group than in the trained embryologist group) the DSMB should use O'Brian-Fleming's sequential boundaries on the positive side. The DSMB should start to look at efficacy data for benefit after 50% of the subjects have completed evaluation of the primary outcome.

For early termination for harm (clinical pregnancy substantially lower in iDA group than in the trained embryologist group) the DSMB should use a Z value of -2.4 and perform first analysis when 20% of subjects have completed evaluation of the primary outcome. All interim analyses will be performed by the DSMB and will be strictly blinded for everybody outside DSMB.

Interim analysis

No other interim analysis will be performed.

7 CHANGES OF ANALYSIS FROM PROTOCOL

No major changes, only more details of the analyses have been added.

8 LISTING OF TABLES AND LISTINGS

Listing of Tables

Table Number	Table Title
14.1.1	Patient Disposition and Data Sets Analysed (ITT Population)
14.1.2	Protocol Deviations Leading to Exclusion from PP Population (ITT Population)
14.1.3.1	Demographics and Baseline Characteristics (ITT Population)
14.1.3.2	Demographics and Baseline Characteristics (FAS Population)

14.1.3.3	Demographics and Baseline Characteristics (PP Population)
14.1.3.4	Main Treatment Variables (ITT Population)
14.1.3.5	Main Treatment Variables (FAS Population)
14.1.3.6	Main Treatment Variables (PP Population)
14.1.4.1	Concomitant Medications (FAS population)
14.2.1.1	Primary Efficacy Analysis (ITT Population)
14.2.1.2	Primary Efficacy Analysis (PP Population)
14.2.1.3	Primary Efficacy Analysis Sensitivity Analyses, adjusted for centre and selected allocation variables (ITT Population)
14.2.1.5	Primary Efficacy Analysis for each centre (ITT Population)
14.2.1.6	Poolability of investigative sites
14.2.2.1	Secondary Efficacy Analysis (FAS Population)
14.2.2.2	Secondary Efficacy Analysis (PP Population)
14.2.3.1	Subgroup Analyses
14.2.3.2	Exploratory Interaction analyses (FAS Population)
14.2.3.3	Results from study: Time evaluation
14.3.2.1	Summary of Adverse Events (Safety Population)
14.3.2.2	All Adverse Events
14.3.2.3	All Adverse Events by maximum reported intensity
14.3.2.4	All Adverse Events by causality
14.3.2.5	All Serious Adverse Events
14.3.2.6	All Adverse Events leading to discontinuation

Listing of graphs

15.1	Figure 1 should be the flow chart of Screening, Eligibility Assessment, randomisation, ITT, PP analysis as well (i.e. CONSORT Flow Chart)
15.2	Percent of Clinical Pregnancy, Biochemical Pregnancy, and Number of sacs with 95% CI for iDAscore and control group. (ITT population). Vertical bar charts with 95% CI
15.3	Unadjusted mean difference with 95% CI between iDAscore and control group regarding Clinical Pregnancy, Biochemical Pregnancy and Number of sacs. (ITT population) Vertical bar charts with 95% CI
15.4	Adjusted mean difference with 95% CI between iDAscore and control group regarding Clinical Pregnancy, Biochemical Pregnancy and Number of sacs. (ITT population) Vertical bar charts with 95% CI
15.5-15.7	The same figures as in 15.2 to 15.4 for Live birth
15.5	Figures regarding distribution of time to evaluation between the two groups and within each group .

	Results from the substudy. (ITT population) Box plots with means.

Listing of Listings

Listing number	Listing Title
16.2.1	Discontinued Patients
16.2.2	Patients with Important Protocol Deviations
16.2.3	Patients Excluded from the Efficacy Analysis
16.2.4.1	Demographics and Baseline Characteristics
16.2.4.2	Main Treatment variables
16.2.4.3	Medical History
16.2.4.3	Prior and Concomitant Medications
16.2.5	Efficacy Variables
16.2.6	Adverse Events

REFERENCES:

1. MacKinnon, J. G., and White, H. (1985). "Some Heteroskedasticity-Consistent Covariance Matrix Estimators with Improved Finite Sample Properties." *Journal of Econometrics* 29:305–325.

2. Guangyong Zou **(**2004**).**" A Modified Poisson Regression Approach to Prospective Studies with Binary Data".
American Journal of Epidemiology, Volume 159, Issue 7, Pages 702–706,

3. A. Dmitrienko, R. B. D'Agostino, (2018). "Multiplicity Considerations in Clinical Trials". New England Journal of Medicine 378;22