PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Embryo selection through non-invasive preimplantation genetic testing with cell-free DNA in spent culture media: A protocol for a multicentre, double-blind, randomized controlled trial
AUTHORS	Huang, Jin; Rong, LI; Zeng, Lin; Hu, Liang; Shi, Juanzi; Cai, Liyi; Yao, Bing; Wang, Xiu-Xia; Xu, Yanwen; Yao, Yuanqing; Wang, Yan; Zhao, Junzhao; Guan, Yichun; Qian, Weiping; Hao, Guimin; Lu, Sijia; Liu, Ping; Qiao, Jie

VERSION 1 – REVIEW

REVIEWER	Nanassy, Laszlo Universitätsklinikum Schleswig-Holstein, Kinderwunschzentrum
REVIEW RETURNED	17-Nov-2021

GENERAL COMMENTS	The introduction lists a number of papers as studies evaluated niPGT-A, but provide no rational of the study. Especially no relevant evaluation of benefits and harms regarding to the existing data on niPGT-A. Also, no clinical data has been discussed regarding the targeted age group even with invasive PGT-A. There is no information on the test that is to be carried out on the spent culture medium. It is stated that there is a 3 grade scale but no previous data has been provided on the specificity and sensitivity of the test. It is not obvious that the ranking based on this test actually cannot decrease the chances of the patients. The reviewer also questions the statements claimed as benefits such as “study results are expected to increase the clinical pregnancy rate, …..”. Also, the statement of “NICS level can be used to provide a reference for future diagnosis and treatment” is not supported with any evidence in the Introduction. This statement has been listed as benefit for either participating or not in the study. The current evidence behind niPGT should be discussed and evaluated in the manuscript for at least make sure that patients receive the correct information to support their decision. Also, a couple of comments. One good benefit of niPGT-A could be to facilitate the slowly changing trend of using IVF instead of ICSI for PGT-A. Is there a specific reason why only ICSI cases are included in the study? It is not clear how morphological ranking works. There is no clear description but the reviewer assumes that authors use the Gardner grading system for blastocysts. Does this description indicates that a day 5 embryo is always better than a day 6? Also, many would argue that a 4AA is worse than a 5AB. If there is an indication in the literature it should be referenced. The volume of each droplet of culture medium cannot exceed 25 uL. Can it be less? If so, what is minimum volume?

REVIEWER	Surdo, Matteo University of Rome Tor Vergata, Virology
REVIEW RETURNED	18-Nov-2021

GENERAL COMMENTS	The protocol presented by Huang Jin, “Embryo selection through non-invasive preimplantation genetic testing with cell-free DNA in spent culture media: A protocol for a multicentre, double-blind, randomized controlled trial”, is well described and rich in information useful to perform the multicenter study. There is a big involvement of China’s hospital centers in order to encourage the use of ni-PGT technique. Minor comments to improve the study protocol: 1. The NICS techniques that will be used in the protocol should be detailed due to the large variability of molecular techniques available. 2. Invasive or non invasive prenatal testing after pregnancy confirmation could help the interpretation of the study results and should be included as analysis. 3. Lines 74-75: As well indicated from the authors, the cost-effectiveness analysis is not included but this parameter could be important for some hospital and IVF centers in order to perform the tests. This information, also in a general point of view, should be included in the protocol. 4. Line 143: The BMI range of females included in the study should not exceed 25. BMI ranges between 25 and 30 are expression of an overweight status and this could be dangerous during

VERSION 1 – AUTHOR RESPONSE

Reviewer 1

Dr. Laszlo Nanassy, Universitätsklinikum Schleswig-Holstein

Comments to the Author:

1. The introduction lists a number of papers as studies evaluated niPGT-A, but provide no rational of the study. Especially no relevant evaluation of benefits and harms regarding to the existing data on niPGT-A.

Response: Thank you for the comment. We agree with the reviewer and have made the necessary changes in the manuscript, line 108-126.

2. Also, no clinical data has been discussed regarding the targeted age group even with invasive PGT-A.

Response: The clinical data in the literatures has been added in the manuscript, line 97-104. Munné et al. (1) conducted a multinational multicenter clinical trial, compared the PGT-A with non-PGT-A. The results showed the ongoing pregnancy rate of PGT-A group aged 35-40 years was statistically increased, but there was no such trend for people <35 years old. Chang et al., (2) comparing 5471 PGT-A cycles with 97069 non-PGT-A cycles, the people age ≥35 in the PGT-A group significantly reduced the rate of miscarriage.

(1).Munné S, et al. Preimplantation genetic testing for aneuploidy versus morphology as selection criteria for single frozen-thawed embryo transfer in good-prognosis patients: a multicenter randomized clinical trial. Fertil Steril 2019;112(6):1071-79.e7. doi: 10.1016/j.fertnstert.2019.07.1346
(2).Chang J, et al. Outcomes of in vitro fertilization with preimplantation genetic diagnosis: an analysis of the United States Assisted Reproductive Technology Surveillance Data, 2011-2012. Fertil Steril 2016;105(2):394-400. doi: 10.1016/j.fertnstert.2015.10.018

3. There is no information on the test that is to be carried out on the spent culture medium. It is stated that there is a 3 grade scale but no previous data has been provided on the specificity and sensitivity of the test.

Response: Because the cfDNA content in the culture medium is very limited and the composition is complexed, the detection of SCM is still in the research stage, there is only one of the participating units made an validation of the accuracy of SCM with a small scale: Chen et al. published an article (1), in which the sensitivity of the NICS grading system according to euploid (A), aneuploid (B), was 87.4%, specificity was 80.3%. And the 3 grade scale was described in a preprint of Chen et al.'s article (2). The grading system had an AUC value of 0.92 and a negative predictive value (NPV) of 0.93.

(1).Chen L, et al. A Non-invasive Chromosome Screening Strategy for Prioritizing in vitro Fertilization Embryos for Implantation Frontiers in Cell and Developmental Biology. 2021, 9. doi:10.3389/fcell.2021.708322.
(2).Chen L, et al. Machine Learning-Guided Noninvasive Embryo Selection for Clinical in Vitro Fertilization Treatment to Avoid Wasting Potentially Qualified Embryos. 2021. doi: 10.21203/rs.3.rs-617438/v1

The AUC value and negative predictive value (NPV) of the grading system have been added in the manuscript, line 275-276.

4. It is not obvious that the ranking based on this test actually cannot decrease the chances of the patients.

Response: In this study, the embryos were transplanted according to the NICS ranking grade, all of the embryos can be transferred, which is different from the aneuploid embryos were not expected to be transferred. Therefore this test actually cannot decrease the chances of the patients. In Chen's preprint article (1), a prospective blinded observational clinical study of the rating system was conducted, which support using the embryo grade system to optimize selection of a single embryo for transfer that will maximize the chance of life birth and avoid the waste of potential qualified embryos.

(1).Chen L, et al. Machine Learning-Guided Noninvasive Embryo Selection for Clinical in Vitro Fertilization Treatment to Avoid Wasting Potentially Qualified Embryos. 2021. doi: 10.21203/rs.3.rs-617438/v1

5. The reviewer also questions the statements claimed as benefits such as “study results are expected to increase the clinical pregnancy rate, …..”. Also, the statement of “NICS level can be used to provide a reference for future diagnosis and treatment” is not supported with any evidence in the Introduction. This statement has been listed as benefit for either participating or not in the study. The current evidence behind niPGT should be discussed and evaluated in the manuscript for at least make sure that patients receive the correct information to support their decision.

Response: As mentioned in the introduction, several studies have found that PGT-A by

trophectoderm (TE) biopsy or whole embryo and SCM has consistency rate ranged from 78.2% to 100% (1-6). However, other research groups have reported relatively low consistency rates, Yin et al. and Ho et al. showed that the concordance was only 32.2% and 56.3% (7-8). At present, several studies have applied non-invasive PGT-A to diverse patient groups to preliminarily evaluate the clinical manifestations of the technique, such as Xu et al., Rubio et al., and Fang et al. used non-invasive PGT-A for patients to improve the clinical outcomes (1,9,4). These data make sure that patients receive the correct information to support their decision.

(1).Xu J, Fang R, Chen L, et al. Noninvasive chromosome screening of human embryos by genome sequencing of embryo culture medium for in vitro fertilization. PNAS, 2016. 113(42): 11907-11912.
(2).Kuznyetsov V, et al. Evaluation of a novel non-invasive preimplantation genetic screening approach. PLoS One 2018;13(5):e0197262.
(3).Rubio C, et al. Multicenter prospective study of concordance between embryonic cell-free DNA and trophectoderm biopsies from 1301 human blastocysts. Am J Obstet Gynecol 2020;223(5):751 e1-51 e13.
(4).Rubio C, et al. Embryonic cell-free DNA versus trophectoderm biopsy for aneuploidy testing: concordance rate and clinical implications. Fertil Steril 2019;112(3):510-19.
(5).Huang L, et al. Noninvasive preimplantation genetic testing for aneuploidy in spent medium may be more reliable than trophectoderm biopsy. Proc Natl Acad Sci U S A 2019;116(28):14105-12.
(6).Jiao J, et al. Minimally invasive preimplantation genetic testing using blastocyst culture medium. Hum Reprod 2019;34(7):1369-79.
(7).Yin B, et al. Validation of preimplantation genetic tests for aneuploidy (PGT-A) with DNA from spent culture media (SCM): concordance assessment and implication. Reproductive biology and endocrinology : RB&E 2021;19, 41.
(8).Ho JR, et al. Pushing the limits of detection: investigation of cell-free DNA for aneuploidy screening in embryos. Fertil Steril 2018;110, 467-475.e462.
(9).Fang R, Yang W, Zhao X, et al. Chromosome screening using culture medium of embryos fertilised in vitro: a pilot clinical study. Journal of translational medicine 2019;17, 73

Also, a couple of comments.
6. One good benefit of niPGT-A could be to facilitate the slowly changing trend of using IVF instead of ICSI for PGT-A. Is there a specific reason why only ICSI cases are included in the study?

Response: Both the ASRM and the ESHRE/PGDIS consensus recommend the use of ICSI for insemination in PGT (1,2). Expert consensus on preimplantation genetic diagnosis/screening in China also pointed out that ICSI insemination is recommended (3). Therefore, in this trail, in order to minimize the impact of paternal sperm on the accuracy of downstream genetic testing, we selected the ICSI patients.

(1).Kokkali G, et al. (2020) ESHRE PGT Consortium and SIG Embryology good practice recommendations for polar body and embryo biopsy for PGT. Human reproduction open 2020(3)p:hoaa020.
(2).Anonymous (2020) Intracytoplasmic sperm injection (ICSI) for non-male factor indications: a committee opinion. Fertil Steril 114(2)p:239-245.
(3).Hefeng H, et al. (2018) Expert Consensus on Preimplantation Genetic Diagnosis/Screening. Chinese Journal of Medical Genetics. doi:10.3760/cma.j.issn.1003-9406.2018.02.001.

7. It is not clear how morphological ranking works. There is no clear description but the reviewer assumes that authors use the Gardner grading system for blastocysts. Does this description indicates that a day 5 embryo is always better than a day 6? Also, many would argue that a 4AA is worse than a 5AB. If there is an indication in the literature it should be referenced.

Response: We followed the Gardner grading system for morphological evaluation. It has been revised in the manuscript, line 281.
According to Gardner grading system, A>B>C (1).
Kovalevsky et al. found that the clinical pregnancy rate and ongoing pregnancy rate of D5 blastocyst were significantly higher than D6 (2). A meta-analysis by Bourdon et al. also showed that patients who transferred D5 embryos got a significantly higher rate of clinical pregnancy and live birth compared to D6 embryos (3). So the D5 embryos are better than D6 embryos.
For morphological grading sequences:
5AA>5AB>5BA>4AA>4AB>4BA>6AA>6AB>6BA>5BB>4BB>6BB>5AC>5BC>
4AC>4BC>6AC>6BC>5CA>5CB>4CA>4CB>6CA>6CB, which is the consensus of the 13 centers participating in this trial.

(1).Gardner DK, Lane M, Stevens J, Schlenker T, & Schoolcraft WB (2000) Blastocyst score affects implantation and pregnancy outcome: towards a single blastocyst transfer. Fertility and sterility 73(6)p:1155-1158.
(2).Kovalevsky G, et al. (2013) Should embryos developing to blastocysts on day 7 be cryopreserved and transferred: an analysis of pregnancy and implantation rates. Fertility and sterility 100(4)p:1008-1012.
(3).Bourdon M, Pocate-Cheriet K, Finet de Bantel A, Grzegorczyk-Martin V, Amar Hoffet A, Arbo E, Poulain M, Santulli P (2019) Day 5 versus Day 6 blastocyst transfers: a systematic review and meta-analysis of clinical outcomes. Hum Reprod 34: 1948-1964

8. The volume of each droplet of culture medium cannot exceed 25 uL. Can it be less? If so, what is minimum volume?

Response: The minimum volume is 20ul according to our internal data. The volume of each droplet of culture medium in this study is 25 μL. About 20ul culture medium was collected into an RNase/DNase-free PCR tube for subsequent testing. It has been revised in the manuscript, line 265-267.

Reviewer: 2
Dr. Matteo Surdo, University of Rome Tor Vergata, Genoma Group Laboratory
Comments to the Author:

The protocol presented by Huang Jin,“Embryo selection through non-invasive preimplantation genetic testing with cell-free DNA in spent culture media: A protocol for a multicentre, double-blind, randomized controlled trial”, is well described and rich in information useful to perform the multicenter study.
There is a big involvement of China’s hospital centers in order to encourage the use of ni-PGT technique.

Response: Thank you for your positive comments on our study.

Minor comments to improve the study protocol:

1. The NICS techniques that will be used in the protocol should be detailed due to the large variability of molecular techniques available.

Response: Thank you for the comment. We agree with the reviewer and have made the necessary changes in the manuscript, line 108-126.

2. Invasive or non invasive prenatal testing after pregnancy confirmation could help the interpretation

of the study results and should be included as analysis.

Response: The patients of this study was 35-42 years old, and in principle, amniocentesis was required. However, this study included 13 reproductive centers across China, and the enrolled patients came from all over the China mainland, and the medical conditions and obstetrics in these places are quite different. Therefore, either the invasive or non invasive prenatal testing was not required, but we will follow up until the delivery. If there are any questions during the follow-up period, corresponding guidance will be given. In this study, we will also collect the products of conception as much as possible for further verification.

3. Lines 74-75: As well indicated from the authors, the cost-effectiveness analysis is not included but this parameter could be important for some hospital and IVF centers in order to perform the tests. This information, also in a general point of view, should be included in the protocol.

Response: Thank you for raising this concern. Cost-effectiveness analysis was not performed because the clinical outcome in our design was the ongoing pregnancy rate of the first transferred cycle and did not involve the cumulative pregnancy rate. At the same time, the time to pregnancy was not collected in this trial, which is difficult to calculate the cost-effectiveness. However, we had collected cumulative pregnancies at unscheduled visits, and cost-effectiveness will be shown in follow-up results.

4. Line 143: The BMI range of females included in the study should not exceed 25. BMI ranges between 25 and 30 are expression of an overweight status and this could be dangerous during pregnancy because increase the risk to develop gestational diabetes.

Response: The BMI set at 18-30kg/m2 was consider to validate the niPGT-A will benefit most of the patients, not just a certain group. Though the patients with BMI range from 25 to 30 had an increased pregnancy risky, but these patients were meet the requirements of ART before entering IVF cycle. If they were not meet the requirements, they will be exclude, and the first exclusion criterion for this study was that “Couples with a contraindication for IVF or ICSI”.
In addition, some polycystic ovarian syndrome (PCOS) patients who may have a BMI over 25, but with appropriate adjustments, IVF can be performed.

Thanks for your consideration.

VERSION 2 – REVIEW

REVIEWER	Kovacs, Peter Kaali Institute
REVIEW RETURNED	09-Feb-2022

GENERAL COMMENTS	Huang et al. revised their study protocol about non-invasive preimplantation genetic testing for embryo selection. This manuscript was already reviewed by two reviewers. As far as I can tell their remarks have been addressed by the authors and the manuscript has been edited accordingly. I have the following additional comments about the paper: 1. Multiple sites have been selected as study centers. Do they use the same culture conditions (gas composition, O2 concertation, culture medium (one step, vs. sequential), culture medium supplementation)? 2. Randomization is performed once the patient has at least two expanded good morphology blastocysts. Previous RCTs on PGT-

	A were criticized for not randomizing patients at cycle start but only when already good quality blastocysts were obtained. This does not necessarily mimic what happens in routine daily practice as there is a significant proportion of patients who do not reach the good morphology blastocyst stage. What percent of the potentially eligible patients will drop-out due to not meeting the blastocyst criteria? Why was the randomization not done at the start of stimulation? Wouldn't that be the real intent-to-treat? 3. On page 18, ln 341-342 they mention that "according to previous work" in similar age women a 38.8% ongoing pregnancy rate can be achieved. There is no reference provided stating whose previous work this is. Is this the average ongoing pregnancy rate in the 13 participating centers for the study age group? When the sample size calculation is described by the authors they mention a 10% drop-out rate? What is considered a drop out in light of the issues raised in points 2 and 4 of these comments? 4. How will they deal with the following issues during niPGT test result interpretation: - No result for amplification failure or non-informative results. These were reported to occur in 2.6% of the cases and 5.6% of the cases respectively by C Rubio et al. (Am J OB Gyn 2020;223:e1-13) - Exogenous (non-embryonic) cf DNA: this was reported to be found in 25% of the cases when looked for according to the systematic review of Brouillet at al (RBMO 2020;40:779-93) - Discrepancies between spent culture medium and invasive PGT-A results. This was reported to occur in 7.4-18.2% of the cases according to the same systematic review. Embryonic mosaicism was brought up as one potential explanation for the discrepancies. Will only those embryos be considered for transfer that provide result and the result is euploid? 5. On page 15, ln 279-281 they state that in the morphological assessment group only the first thawed embryo will be assessed. It seems that this thawing is however completely random. They probably assess embryo morphology and score the embryos before they are actually vitrified. Wouldn't it make sense to thaw the embryo that had the highest score for morphology at the time of vitrification? 6. The authors should discuss what the intention-to-treat and per protocol populations are since patients are randomized once they already have two or more expanded blastocysts. It would be nice to have a reliable, non-invasive test that can identify euploid embryos. Therefore, I support the study idea and look forward to the results. Certain issues should be explained or clarified about the study methods raised in the above comments.

VERSION 2 – AUTHOR RESPONSE

Reviewer 3

Dr. Peter Kovacs, Kaali Institute

Comments to the Author:

Huang et al. revised their study protocol about non-invasive preimplantation genetic testing for embryo selection.

This manuscript was already reviewed by two reviewers. As far as I can tell their remarks have been addressed by the authors and the manuscript has been edited accordingly.

I have the following additional comments about the paper:

1. Multiple sites have been selected as study centers. Do they use the same culture conditions (gas composition, O2 concertation, culture medium (one step, vs. sequential), culture medium supplementation)?

Response: Thank you for this comment. Blastocyst culture is done with sequential media in three-gas system in all centers. Taking into account the existing processes of the multiple centers, the culture medium is from different manufacturers (Vitro life, Cook, SAGE and Quinns). It has been revised in the line 282-285 of the manuscript.

2. Randomization is performed once the patient has at least two expanded good morphology blastocysts. Previous RCTs on PGT-A were criticized for not randomizing patients at cycle start but only when already good quality blastocysts were obtained. This does not necessarily mimic what happens in routine daily practice as there is a significant proportion of patients who do not reach the good morphology blastocyst stage. What percent of the potentially eligible patients will drop-out due to not meeting the blastocyst criteria? Why was the randomization not done at the start of stimulation? Wouldn't that be the real intent-to-treat?

Response: Thank you for this comment. In fact, these patients were included in this trial because the purpose of this study is focus on the clinical value of niPGT-A as a new effective indicator to evaluate embryo. The niPGT-A is an embryo preferred method. The blastocysts were classified into three grades from A to C according to their euploid probabilities predicted by the niPGT-A results. A single blastocyst will be thawed and transferred in the preference order of A>B>C. If the patients has only 1 blastocyst, the niPGT-A result will not be able to provide a reference for embryo selection. This is different from the RCTs on PGT-A, which divides the embryos into euploidy and aneuploidy, and the euploid embryo was transferred. Therefore, patients with two good quality blastocysts were included in this study.

3. On page 18, ln 341-342 they mention that "according to previous work" in similar age women a 38.8% ongoing pregnancy rate can be achieved. There is no reference provided stating whose previous work this is. Is this the average ongoing pregnancy rate in the 13 participating centers for the study age group? When the sample size calculation is described by the authors they mention a 10% drop-out rate? What is considered a drop out in light of the issues raised in points 2 and 4 of these comments?

Response: Yes, 38.8% ongoing pregnancy rate is the average ongoing pregnancy rate in the 13 participating centers for the study age group. It has been added in the manuscript, line 362-363.

Considering fully completed the clinical trial with a relatively generous estimation method, we calculated the sample size according to the 10% dropout rate. About the final accurate dropout rate, we will calculate from the final data after the end of the trial.

Determination of dropout: All patients who signed the informed consent form have the right to withdraw their consent and quit at any stage of the trial. After randomization, patients who do not meet the protocol are considered to be dropped out.

Reasons for dropout:

Researchers decide to drop out:

1) Serious adverse events occur, and the patients are not suitable to continue the study according to the researchers' judgment;

2) Before the blastocyst transfer, the patients have other diseases or special physiological changes and are not suitable to continue;

3) Patients who do not meet the trial protocol;

4) In case of emergencies during the study, the blindness of the patients are broken.

Patients withdraw:

1) For whatever reason, the patients are unwilling to continue the trial process or withdraw the consent;

2) Although the patients do not withdraw the consent, but they no longer receive visits and follow-up.

The determination of dropout has been added in the manuscript, line 186-203.

4. How will they deal with the following issues during niPGT test result interpretation:

- No result for amplification failure or non-informative results. These were reported to occur in 2.6% of the cases and 5.6% of the cases respectively by C Rubio et al. (Am J OB Gyn 2020;223:e1-13)

- Exogenous (non-embryonic) cf DNA: this was reported to be found in 25% of the cases when looked for according to the systematic review of Brouillet at al (RBMO 2020;40:779-93)

- Discrepancies between spent culture medium and invasive PGT-A results. This was reported to occur in 7.4-18.2% of the cases according to the same systematic review. Embryonic mosaicism was brought up as one potential explanation for the discrepancies.

Will only those embryos be considered for transfer that provide result and the result is euploid?

Response:

(1) Embryos with no result for amplification failure or non-informative results are graded B. According to the probability of euploidy, we divided the embryos into three grades. A single blastocyst will be thawed and transferred in the preference order of A>B>C. As the clinical results shown in Chen's preprint article (DOI: 10.21203/rs.3.rs-617438/v1, now accept by RBMO), the live birth rate in A- versus C-grade embryos was 50.4% versus 27.1% (p=0.006) and B- versus C-grade embryos was

45.3% versus 27.1% (p=0.022) ; the miscarriage rate in A- versus C-grade embryos was 15.9% versus 33.3% (p=0.026) and B- versus C-grade embryos was 14.3% versus 33.3% (p=0.021). In addition, the results of amplification failure or non-informative may be related to the low content of cfDNA in the media. According to the report of Magli et al. (Fertility and sterility 2019; 111(1)p:77-85.), transferring an embryo with successful blastocoel fluid amplification led to a clinical pregnancy rate of only 37% and an ongoing pregnancy rate of 18%, while transferring an embryo with blastocoel fluid amplification failure resulted in a clinical pregnancy rate of 77% and an ongoing pregnancy rate of 70%. It means that embryos with failed amplification or non-informative may also be transferred, but because the karyotype is unknown, they cannot be graded A, so embryos with failed amplification or non-informative are graded B.

(2) About the exogenous (non-embryonic) cfDNA, the article (J Clin Invest. 2021;131(12):e146051) also showed that the components in the culture are relatively complex, and there is the possibility of maternal DNA contamination. In order to avoid exogenous contamination, our latest article (JoVE 2021.(175)) standardized the sampling method of the culture medium. In this trial, the embryos will be replaced with the culture medium on D3, and the granulosa cells will be removed with an egg stripper again. Then the culture medium will be replaced in the afternoon of D4. These operations can effectively remove exogenous cfDNA contamination, and also ensures the accuracy of niPGT-A results. Our purpose is not to diagnose whether the embryos are euploid, but sort the embryos according to the probability of euploidy. This detection method is expected to be another criteria for embryo evaluation.

(3) Indeed, some studies have reported that there are discrepancies between spent culture medium and invasive PGT-A results. Due to embryo mosaicism, the results of TE sometimes cannot reflect the real condition of ICM, while cfDNA in the SCM is derived from whole embryo, which may better reflect the results of whole embryo. We considered the mosaicism in the grading system and gave a prediction of the probability of euploidy for each embryo, suggesting that embryos with a higher probability of euploidy should be preferentially transferred.

(4) No, in fact, all embryos are considered for transfer, but there is a priority. The grading system is judged according to the probability of euploidy. The euploid probability of grade A embryos is the highest, so the priority of transfer is the highest. But if there is no grade A embryos, we also recommend grade B embryos for transfer, then grade C embryos with full informed consent. If the patient has multiple embryos of the same grade, the embryo with the highest morphological grade will be selected for transfer.

5. On page 15, ln 279-281 they state that in the morphological assessment group only the first thawed embryo will be assessed. It seems that this thawing is however completely random. They probably assess embryo morphology and score the embryos before they are actually vitrified. Wouldn't it make sense to thaw the embryo that had the highest score for morphology at the time of vitrification?

Response:

Thank you for the comment. In this study, all embryos were morphologically evaluated before freezing. The morphological group thawed the single embryo with the highest morphological rating (for multiple embryos with the same rating, one was randomly selected and thawed). For the single thawed embryo, we will record the morphological rating after thawing for subsequent analysis.

In clinical operations, it is impossible to thaw and perform morphological scoring of all embryos. So we refer to the morphological scores of embryos before cryopreservation. Among them, the morphological group thawed and transfer the embryo with the highest morphological score, while the

niPGT-A group based on the ABC grade. In blastocyst with the same grade, blastocysts with a higher morphology grade would be preferentially transferred.

6. The authors should discuss what the intention-to-treat and per protocol populations are since patients are randomized once they already have two or more expanded blastocysts.

Response: Thank you for the comment. The intention-to-treat (ITT) populations refer to all randomized patients who experienced at least one intervention and had post-intervention evaluation data. The per protocol (PPS) populations refer to those who meet at least the following criteria: (1) Meet the inclusion criteria and follow the protocol; (2) Complete all planned visits; (3) No drugs or treatments that may affect the evaluation of efficacy were used during the trial. The part of the content was added in the line 374-380 of the manuscript.

VERSION 3 – REVIEW

REVIEWER	Kovacs, Peter Kaali Institute
REVIEW RETURNED	06-Apr-2022

GENERAL COMMENTS	Huang et. revised their study protocol and made certain changes in their manuscript. I have the following comments about the revised RCT protocol: 1. I have not received a point-by-point response to the issues raised. This may not have been a requirement though it certainly would have been easier to see how they addressed the questions raised. In addition, some of the points raised by me were not answered; no changes were implemented and therefore seems that these have been ignored during the revision of the protocol. 2. The authors have now included that similar laboratory techniques are used in the 13 study centers. 3. There is no explanation provided why the randomization was at the blastocyst stage for those who had at least two good morphology blastocysts available and why not at the start of stimulation (i.e. start of treatment). The reference according to which the blastocysts are scored ("morphology is greater than 4BC/4CB..") should be included in line 173 where first mentioned. 4. It is now explained based on what pregnancy rate was the power analysis performed. 5. Exclusion criteria #2 is: "those with preimplantation genetic testing". Half of the participants will undergo preimplantation genetic testing (from spend culture medium) though not through trophectoderm biopsy. Do they mean those who elect to undergo trophectoderm biopsy will be excluded or do they mean that those who undergo preimplantation genetic testing for monogenic disorders or structural chromosome defect will be excluded? 6. The drop out definitions are not clear. Patients are randomized at the blastocyst stage as one of the criteria is to have two good quality blastocysts. Therefore, they have completed the entire IVF treatment minus the embryo transfer. By drop out, do they mean those who do not come back for the frozen embryos to be transferred for medical or personal reasons? Drop out criteria #3: "those who do not meet the study protocol". How could this

	happen? The patients must meet the inclusion-exclusion criteria by the time they get to the blastocyst stage to be considered for randomization. How do they handle those cases when the embryo does not survive the thawing? How do they manage those cases when there is no genetic test result for the reasons listed in point 4 of my initial review? They need to be a lot more precise when the drop out reasons are listed. 7. It is still not explained which one is the “first thawed embryo” in the morphological group. Is this the one that was considered the best at the time of cryopreservation or is this first thawed embryo chosen completely randomly? 8. The various FET protocols should be described briefly. 9. I still miss the correct explanation of the ITT and PP populations. According to the current description the ITT population is the population that had at least one intervention. Since the randomization (study entry) is done at the end of the IVF treatment at blastocyst stage, the next “intervention” these patients will undergo is the FET. Do they mean any intervention as part of the FET (e.g.: ultrasound)? How do the drop outs affect the ITT and PP populations? What happens to the patient who has two good blastocysts cryopreserved in the morphological group but the embryos do not survive the thawing? Is this patient included in the ITT but not the PP population? What happens to the patient who received no niPGT result for technical problems? Again, is she included in the ITT population but not in the PP population? I still look forward to the study result but the study protocol to be accepted for publication the above questions should be properly answered.

VERSION 3 – AUTHOR RESPONSE

Response to reviewer’s Comments:

1. I have not received a point-by-point response to the issues raised. This may not have been a requirement though it certainly would have been easier to see how they addressed the questions raised. In addition, some of the points raised by me were not answered; no changes were implemented and therefore seems that these have been ignored during the revision of the protocol.

Response: Previously, we replied point by point in the reply letter, but some details were not confirmed whether they need to be added to the text. This time we have added all the content that can be added, replied in the marked manuscript and reply letter. The following is the content of the last reply letter. The red font is the newly added (If the red font did not show in this letter, please see the the marked manuscript).

1. Multiple sites have been selected as study centers. Do they use the same culture conditions (gas composition, O2 concertation, culture medium (one step, vs. sequential), culture medium supplementation)?

Response: Thank you for this comment. Blastocyst culture is done with sequential media in three-gas system in all centers. Taking into account the existing processes of the multiple centers, the culture medium is from different manufacturers (Vitro life, Cook, SAGE and Quinns). It has been revised in the line 304-307 of the manuscript.

2. Randomization is performed once the patient has at least two expanded good morphology blastocysts. Previous RCTs on PGT-A were criticized for not randomizing patients at cycle start but only when already good quality blastocysts were obtained. This does not necessarily mimic what happens in routine daily practice as there is a significant proportion of patients who do not reach the good morphology blastocyst stage. What percent of the potentially eligible patients will drop-out due to not meeting the blastocyst criteria? Why was the randomization not done at the start of stimulation? Wouldn't that be the real intent-to-treat?

Response: Thank you for this comment. In fact, these patients were included in this trial because the purpose of this study is focus on the clinical value of niPGT-A as a new effective indicator to evaluate embryo. The niPGT-A is an embryo preferred method. The blastocysts were classified into three grades from A to C according to their euploid probabilities predicted by the niPGT-A results. A single blastocyst will be thawed and transferred in the preference order of A>B>C. If the patients has only 1 blastocyst, the niPGT-A result will not be able to provide a reference for embryo selection. This is different from the RCTs on PGT-A, which divides the embryos into euploidy and aneuploidy, and the euploid embryo was transferred. Therefore, patients with two good quality blastocysts were included in this study.
Add response: It has been revised in the line 327-330 of the marked manuscript.

3. On page 18, ln 341-342 they mention that "according to previous work" in similar age women a 38.8% ongoing pregnancy rate can be achieved. There is no reference provided stating whose previous work this is. Is this the average ongoing pregnancy rate in the 13 participating centers for the study age group? When the sample size calculation is described by the authors they mention a 10% drop-out rate? What is considered a drop out in light of the issues raised in points 2 and 4 of these comments?

Response: Yes, 38.8% ongoing pregnancy rate is the average ongoing pregnancy rate in the 13 participating centers for the study age group. It has been added in the manuscript, line 414-415. Considering fully completed the clinical trial with a relatively generous estimation method, we calculated the sample size according to the 10% dropout rate. About the final accurate dropout rate, we will calculate from the final data after the end of the trial.
Add response: It has been added in the manuscript, line 422-424.
Determination of dropout: All patients who signed the informed consent form have the right to withdraw their consent and quit at any stage of the trial.
Reasons for dropout:
Researchers decide to drop out:
1) Serious adverse events occur, and the patients are not suitable to continue the study according to the researchers' judgment;
2) Before the blastocyst transfer, the patients have other diseases or special physiological changes and are not suitable to continue;
3) Patients who don't have the first blastocyst transfer within 6 months after oocyte retrieval;
4) Patients who don't complete the first transfer in the preferred order due to thawing failure.
5) In case of emergencies during the study, the blindness of the patients are broken.
Patients withdraw:
1) For whatever reason, the patients are unwilling to continue the trial process or withdraw the consent;
2) Although the patients do not withdraw the consent, but they no longer receive visits and follow-up.
Add response: The determination of dropout has been revised based on the latest suggestions in the marked manuscript, line 206-225.

4. How will they deal with the following issues during niPGT test result interpretation:
- No result for amplification failure or non-informative results. These were reported to occur in 2.6% of

the cases and 5.6% of the cases respectively by C Rubio et al. (Am J OB Gyn 2020;223:e1-13)
- Exogenous (non-embryonic) cf DNA: this was reported to be found in 25% of the cases when looked for according to the systematic review of Brouillet at al (RBMO 2020;40:779-93)
- Discrepancies between spent culture medium and invasive PGT-A results. This was reported to occur in 7.4-18.2% of the cases according to the same systematic review. Embryonic mosaicism was brought up as one potential explanation for the discrepancies.
Will only those embryos be considered for transfer that provide result and the result is euploid?

Response:
(1) Embryos with no result for amplification failure or non-informative results are graded B. According to the probability of euploidy, we divided the embryos into three grades. A single blastocyst will be thawed and transferred in the preference order of A>B>C. As the clinical results shown in Chen's preprint article (https://www.sciencedirect.com/science/article/abs/pii/S1472648322001420, now accept by RBMO), the live birth rate in A- versus C-grade embryos was 50.4% versus 27.1% (p=0.006) and B- versus C-grade embryos was 45.3% versus 27.1% (p=0.022) ; the miscarriage rate in A- versus C-grade embryos was 15.9% versus 33.3% (p=0.026) and B- versus C-grade embryos was 14.3% versus 33.3% (p=0.021). In addition, the results of amplification failure or non-informative may be related to the low content of cfDNA in the media. According to the report of Magli et al. (Fertility and sterility 2019; 111(1)p:77-85.), transferring an embryo with successful blastocoel fluid amplification led to a clinical pregnancy rate of only 37% and an ongoing pregnancy rate of 18%, while transferring an embryo with blastocoel fluid amplification failure resulted in a clinical pregnancy rate of 77% and an ongoing pregnancy rate of 70%. It means that embryos with failed amplification or non-informative may also be transferred, but because the karyotype is unknown, they cannot be graded A, so embryos with failed amplification or non-informative are graded B.
Add response: It has been added in the manuscript, line 323-324.

(2) About the exogenous (non-embryonic) cfDNA, the article (J Clin Invest. 2021;131(12):e146051) also showed that the components in the culture are relatively complex, and there is the possibility of maternal DNA contamination. In order to avoid exogenous contamination, our latest article (JoVE 2021.(175)) standardized the sampling method of the culture medium. In this trial, the embryos will be replaced with the culture medium on D3, and the granulosa cells will be removed with an egg stripper again. Then the culture medium will be replaced in the afternoon of D4. These operations can effectively remove exogenous cfDNA contamination, and also ensures the accuracy of niPGT-A results. Our purpose is not to diagnose whether the embryos are euploid, but sort the embryos according to the probability of euploidy. This detection method is expected to be another criteria for embryo evaluation.
Add response: It has been added in the manuscript, line 308-310.

(3) Indeed, some studies have reported that there are discrepancies between spent culture medium and invasive PGT-A results. Due to embryo mosaicism, the results of TE sometimes cannot reflect the real condition of ICM, while cfDNA in the SCM is derived from whole embryo, which may better reflect the results of whole embryo. We considered the mosaicism in the grading system and gave a prediction of the probability of euploidy for each embryo, suggesting that embryos with a higher probability of euploidy should be preferentially transferred.
Add response: It has been added in the manuscript, line 320-327.

(4) No, in fact, all embryos are considered for transfer, but there is a priority. The grading system is judged according to the probability of euploidy. The euploid probability of grade A embryos is the highest, so the priority of transfer is the highest. But if there is no grade A embryos, we also recommend grade B embryos for transfer, then grade C embryos. If the patient has multiple embryos of the same grade, the embryo with the highest morphological grade will be selected for transfer.

Add response: It has been added in the manuscript, line 323-326.

5. On page 15, ln 279-281 they state that in the morphological assessment group only the first thawed embryo will be assessed. It seems that this thawing is however completely random. They probably assess embryo morphology and score the embryos before they are actually vitrified. Wouldn't it make sense to thaw the embryo that had the highest score for morphology at the time of vitrification?

Response:
Thank you for the comment. In this study, all embryos were morphologically evaluated before freezing. The morphological group thawed the single embryo with the highest morphological rating, which is evaluated before freezing. If a patient has multiple embryos with the highest rating, embryologists selected and thawed the best embryo between the highest rating according to the center standards. For the single thawed embryo, we will record the morphological rating after thawing for subsequent analysis.
In clinical operations, it is impossible to thaw and perform morphological scoring of all embryos. So we refer to the morphological scores of embryos before cryopreservation. Among them, the morphological group thawed and transfer the embryo with the highest morphological score. The niPGT-A group based on the ABC grade, in blastocyst with the same NICS grade, blastocysts with a higher morphology grade would be preferentially transferred.
Add response: The content was revised in the manuscript, line 313-314, 334-341.

6. The authors should discuss what the intention-to-treat and per protocol populations are since patients are randomized once they already have two or more expanded blastocysts.

Response: Thank you for the comment. The intention-to-treat (ITT) populations refer to all randomized patients who experienced at least one intervention and had post-intervention evaluation data. The per protocol (PPS) populations refer to those who meet at least the following criteria: (1) Meet the inclusion criteria and follow the protocol; (2) Complete all planned visits; (3) No drugs or treatments that may affect the evaluation of efficacy were used during the trial. The part of the content was added in the line 428-436 of the marked manuscript.

2. The authors have now included that similar laboratory techniques are used in the 13 study centers.

Response: Yes, the content was added in the line 304-307 of the marked manuscript.

3. There is no explanation provided why the randomization was at the blastocyst stage for those who had at least two good morphology blastocysts available and why not at the start of stimulation (i.e. start of treatment). The reference according to which the blastocysts are scored ("morphology is greater than 4BC/4CB..") should be included in line 173 where first mentioned.

Response: (1). As stated in the last reply letter, in fact, these patients were included in this trial because the purpose of this study is focus on the clinical value of niPGT-A as a new effective indicator to evaluate embryo. The niPGT-A is an embryo preferred method. The blastocysts were classified into three grades from A to C according to their euploid probabilities predicted by the niPGT-A results (Chen, et al. RBMO, https://www.sciencedirect.com/science/article/abs/pii/S1472648322001420). A single blastocyst will be thawed and transferred in the preference order of A>B>C. If the patients has only 1 blastocyst, the niPGT-A result will not be able to provide a reference for embryo selection. If the randomization was at the start of stimulation, it is not sure to confirm whether the patient has at least 2 good blastocysts, and NICS cannot play a preferred role. Therefore, patients with two good quality blastocysts were included in this study. The content was added in the line 320-330 of the marked manuscript.

(2). The reference according to which the blastocysts are scored ("morphology is greater than 4BC/4CB..") has be included in the line 185.

4. It is now explained based on what pregnancy rate was the power analysis performed.

Response: Yes, the content was added in the line 414-415 of the marked manuscript.

5. Exclusion criteria #2 is: "those with preimplantation genetic testing". Half of the participants will undergo preimplantation genetic testing (from spend culture medium) though not through trophectoderm biopsy. Do they mean those who elect to undergo trophectoderm biopsy will be excluded or do they mean that those who undergo preimplantation genetic testing for monogenic disorders or structural chromosome defect will be excluded?

Response: The preimplantation genetic testing in Exclusion criteria #2, including preimplantation genetic testing for aneuploidy (PGT-A), preimplantation genetic testing for monogenic disorders (PGT-M) or structural chromosome defect (PGT-SR). The content was added in the line 195-197 of the marked manuscript.

6. The drop out definitions are not clear. Patients are randomized at the blastocyst stage as one of the criteria is to have two good quality blastocysts. Therefore, they have completed the entire IVF treatment minus the embryo transfer. By drop out, do they mean those who do not come back for the frozen embryos to be transferred for medical or personal reasons?

Response: (1). To clarify the definition of drop out, in the line 215-218, Drop out criteria #3 is "Patients who don't have the first blastocyst transfer within 6 months after oocyte retrieval". Criteria #4 is added "Patients who don't complete the first transfer in the preferred order due to thawing failure".
(2). Yes, those who do not come back within 6 months after oocyte retrieval for the frozen embryos to be transferred for medical or personal reasons, was included the drop out criteria #3.

Drop out criteria #3: "those who do not meet the study protocol". How could this happen?

Response: Those patients meet the Inclusion and Exclusion criteria when randomization, but due to some factors, the first embryo transfer was not performed according to the protocol within 6 months after oocyte retrieval. Or patients who don't complete the first transfer in the preferred order due to thawing failure. See the line 215-218.

The patients must meet the inclusion-exclusion criteria by the time they get to the blastocyst stage to be considered for randomization. How do they handle those cases when the embryo does not survive the thawing? How do they manage those cases when there is no genetic test result for the reasons listed in point 4 of my initial review? They need to be a lot more precise when the drop out reasons are listed.

Response: Those cases when the embryo does not survive the thawing, were drop out. The content was in line 217-218.
Those cases when there is no genetic test result, were the B grade. The content was in line 323-324.

7. It is still not explained which one is the "first thawed embryo" in the morphological group. Is this the one that was considered the best at the time of cryopreservation or is this first thawed embryo chosen completely randomly?

Response: All embryos were morphologically evaluated before freezing. The morphological group thawed the single embryo with the highest morphological rating. In the first transfer cycle,

embryologists selected and thawed the highest rating embryo. If multiple embryos were the highest rating, the embryologists select the best embryo between the highest rating embryos according to the center standards. The content was added in the line 313-314, 334-341.

8. The various FET protocols should be described briefly.

Response: FET protocols were performed according to the standard of each centre, including natural cycle, hyperstimulation cycle. Endometrial preparation includes natural cycle and artificial cycle. Generally, the window of implantation is on the 7th day after the peak of luteinizing hormone (LH+7) in the natural cycle, or the 5th day when progesterone (P+5) is added in the artificial cycle. Embryo transfer was performed during the window. The details are implemented in accordance with the standards of each centre. (added to lines 356-363)

9. I still miss the correct explanation of the ITT and PP populations. According to the current description the ITT population is the population that had at least one intervention. Since the randomization (study entry) is done at the end of the IVF treatment at blastocyst stage, the next "intervention" these patients will undergo is the FET. Do they mean any intervention as part of the FET (e.g.: ultrasound)?

Response: (1).The intervention is different embryo selection methods, the ni-PGT group chooses the one with NICS highest rating, and the morphological group chooses the one with highest morphological rating. (2).The part of the FET (e.g.: ultrasound), which is a clinical routine operation, can't effect the embryo selection and is not within the scope of our defined intervention. We added this part in the line 315-318.

How do the drop outs affect the ITT and PP populations? What happens to the patient who has two good blastocysts cryopreserved in the morphological group but the embryos do not survive the thawing? Is this patient included in the ITT but not the PP population? What happens to the patient who received no niPGT result for technical problems? Again, is she included in the ITT population but not in the PP population?

Response: Dropout may have an impact on the ITT and PP datasets, which can be used for statistical analysis. Considering fully completed the clinical trial with a relatively generous estimation method, we calculated the sample size according to the 10% dropout rate, line 422-424。
Yes, those cases when the embryo does not survive the thawing for the first transfer with within 6 months after oocyte retrieval, were included the ITT but not PP population. The content was in line 434-436, 217-218.
No, those cases when there is no genetic test result, were the B grade and included the ITT and PP population. The content was in line 323-324.

VERSION 4 – REVIEW

REVIEWER	Kovacs, Peter Kaali Institute
REVIEW RETURNED	02-May-2022

GENERAL COMMENTS	Huang et. further revised their study protocol. The authors have now provided a point-by-point response to the issues raised in my previous review. Most questions have been answered properly and when possible the changes were included in the revised manuscript.

	I still have some minor issues with the drop-out and intention-to-treat vs. per protocol populations. My understanding is that the patients sign their consent form before the stimulation is started. This specifically should be said in the paper as it is not mentioned at what stage they sign (at BC stage or before start of stimulation or anytime in between?). Based on table 1 it seems that the consent form is signed before the stimulation is started. In their response to the reviewer they say that the patient may withdraw her consent any time after signing the form. Therefore my understanding is that some patient will “drop out” before they get to the BC stage (before they are even randomized). When they discuss “determination of dropout” they seem to talk about events that could happen only once the patient has reached the two BC stage and has been randomized. It should specifically stated that drop-outs are only considered after the randomization and those who withdraw their consent, have the stimulation cancelled, have poor fertilization, have suboptimal embryo development or whose embryos do not reach the BC stage (this all should be more than 10%) are not drop-outs. This should explain who the intention-to-treat and the per protocol populations are too. It also should be explained why those embryos that end up with no result for amplification failure or non-informative results are graded as B medium category and why not A or C or a special, fourth category, If these issues are answered and included in the paper I would support its publication.

VERSION 4 – AUTHOR RESPONSE

Response to reviewer’s Comments:

1.I still have some minor issues with the drop-out and intention-to-treat vs. per protocol populations. My understanding is that the patients sign their consent form before the stimulation is started. This specifically should be said in the paper as it is not mentioned at what stage they sign (at BC stage or before start of stimulation or anytime in between?). Based on table 1 it seems that the consent form is signed before the stimulation is started. In their response to the reviewer they say that the patient may withdraw her consent any time after signing the form. Therefore my understanding is that some patient will “drop out” before they get to the BC stage (before they are even randomized). When they discuss “determination of dropout” they seem to talk about events that could happen only once the patient has reached the two BC stage and has been randomized. It should specifically stated that drop-outs are only considered after the randomization and those who withdraw their consent, have the stimulation cancelled, have poor fertilization, have suboptimal embryo development or whose embryos do not reach the BC stage (this all should be more than 10%) are not drop-outs. This should explain who the intention-to-treat and the per protocol populations are too.

Response :
Couples who agree to participate will be asked to sign the consent form on or before the day of oocyte retrieval. The content was added in the manuscript, line 225.
After signing the informed consent form and before randomization, during this period, if the patient withdraws the informed consent, the patient will not be included in the trial and can’t be regarded as drop-outs. We added to the determination of dropout what the reviewer suggested: Drop-outs are only considered after the randomization and those who withdraw their consent, have the stimulation cancelled, have poor fertilization, have suboptimal embryo development or whose embryos do not reach the 4BC/4CB stage are not drop-outs. The content was added in the manuscript, line 198-201.

Regarding the ITT and PP population, the content was in the line 426-428.

2.It also should be explained why those embryos that end up with no result for amplification failure or non-informative results are graded as B medium category and why not A or C or a special, fourth category.

Response:
In the pilot study for this trail (https://www.sciencedirect.com/science/article/abs/pii/S1472648322001420, Reproductive BioMedicine Online), the embryos with no result for amplification failure or non-informative results were included in grade B. The clinical outcomes of transferring B-grade embryos were worse than that of A-grade embryos and better than that of C-grade embryos. Moreover, no or non- informative results were neither euploidy nor aneuploidy, so these embryos were not suitable to be graded as A or C. Therefore, embryos with no result for amplification failure or non-informative results are graded B. The content was added in the line 316-323 of marked manuscript.
Magli's research showed that the ongoing pregnancy rate was 68% in the group with failed blastocyst fluid amplification, and 31.5% in the group with successful blastocyst fluid amplification when followed the clinical outcome of 53 TE-euploid blastocysts (Fertility and sterility 2019; 111(1)p:77-85.). The results indicated that embryos with failed cfDNA amplification also have high transfer potential, and it is not recommended to completely abandon transfer.
In addition, when this clinical trial is completed, we will analyze no result for amplification failure or non-informative results in grade B separately.

We would like to thank the referee again for taking the time to review our manuscript.