

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)	Observational retrospective study of U.S. national utilization patterns and live birth rates for various ovarian stimulation protocols for in vitro fertilization
AUTHORS	Kushnir, Vitaly; Darmon, Sarah; Barad, David; Gleicher, No

VERSION 1 – REVIEW

REVIEWER	Dr Panagiotis Drakopoulos Center for reproductive medicine UZ Brussel Belgium
REVIEW RETURNED	11-Apr-2018

GENERAL COMMENTS	<p>I read with great interest the observational retrospective study by Kushir et al (using the SART database), aiming to evaluate the different protocols used for ovarian stimulation in IVF. It is a hot topic. The authors concluded that conventional stimulation results in significantly higher LBR compared to other protocols (minimal, natural cycle, IVM)</p> <p>Major remarks</p> <p>-A severe limitation of the analysis is the absence of standardized definitions for each protocol (i.e the definition of minimal stimulation is different from the common definition of mild stimulation) and the inability to adjust for relevant confounders. Therefore, no firm conclusion can be drawn, but only speculations....with a risk of bias</p> <p>-Another important limitation is that LBR, defined as the 1st fresh or frozen live birth (in freeze all cycles) is not a relevant endpoint in modern ART, if cumulative results cannot be provided. For each infertile patient, cumulative results (fresh+frozen live birth rates after use of all available embryos) represent the most meaningful outcome. This should be the primary outcome in case we aim to evaluate different stimulation protocols, in the future.</p> <p>-Furthermore, the number of cycles (with minimal/natural/IVF) is significantly smaller compared to conventional ovarian stimulation cycles... making the study groups significantly imbalanced. COS cycles account for > 90% of the data. The authors should comment on it, taking also into account the inability to adjust for confounders. As also reported by the authors, the vast majority of minimal/natural IVF protocols were used in 2 US centers, increasing the risk of selection bias.</p> <p>Minor remarks</p>
-------------------------	---

	<p>-Figure 1: Please provide the number of cycles for each protocol and age strata. I would also suggest to report the results of LBR per protocol and age, in a supplementary Table</p> <p>- Further statistical analyses should also be done for comparisons between the other protocols (minimal vs natural etc)</p> <p>-Why repeated cycles could not be determined in the analysis, so as to be able to adjust for it? Is this a limitation of the SART database?</p> <p>-Could other causes of infertility be reported?</p> <p>-Please comment on the indication of IVM protocols. Do they include exclusively PCOS patients or also patients with fertilization failures?</p> <p>- Authors should be more conservative with their conclusions in general. The fact the mild stimulation could be an effective option for young good prognosis women in term of fresh live birth rates (not cumulative) is a widely accepted concept (based on meta-analyses). From a clinical point of view, please comment on why COS should be a better choice in women > 42 compared to minimal IVF? I don t think that the analyzed data (with the aforementioned limitations) is robust enough to draw this conclusion</p> <p>- the words: observational retrospective should be added in the title</p> <p>-line 56: please add ref 18</p> <p>- there is an extra asterisk in Figure 1 for age strata 41-42. To delete please.</p>
--	---

REVIEWER	Yanping Kuang Shanghai Ninth People's Hospital, Shanghai Jiaotong University School of Medicine
REVIEW RETURNED	17-Apr-2018

GENERAL COMMENTS	<p>This is a retrospective cohort study using the aggregate data of U.S. fertility centers for autologous IVF cycles performed during 2014 and 2015 and presented the utilization patterns and live birth rates differ at varying ages with various ovarian stimulation protocols. Although the current data provide valuable information for IVF practice, there are some flaws in the current version:</p> <ol style="list-style-type: none"> 1. The results presented the limited information including only age, stimulation protocols, whether DOR or not, number of transferred embryos and live birth outcomes which limited the value of this article. More demographic data and cycle characteristics are beneficial to explain the current results. For example, the differences in the definition of stimulation protocols, the characteristics of patients, infertility factors, the treatment or lab techniques among different fertility centers lead to a significant bias. The further stratified analysis and/or the multivariable logistic regression could be useful for live birth rate controlling for these confounding variables. Without these supporting data, the results can not be generalized into the clinical practices, and the result explanations should be with more cautions in the discussion. 2. The live birth rate came from the one ovarian stimulation cycle
-------------------------	--

	<p>and fresh embryo transfer, or once stimulation and all subsequent transfer (fresh and frozen embryos) should be described in methods. The difference in live birth outcomes not only come from the stimulation protocols, but also had some confused factors (age, ovarian reserve, gravity, infertility cause etc). The data about the proportion of cancel cycles or no viable embryo transfer are not available.</p> <p>3. The definition of conventional ovarian stimulation and the minimal ovarian stimulation is not clear. For example, ‘the administration of gonadotropins for 8-10 days to recruit multiple mature eggs’ and ‘use less dose of gonadotropins than conventional ovarian stimulation’, the cutoff dose of gonadotropins between conventional ovarian stimulation and minimal stimulation maybe described as a specific range, or provide more detailed data to make it clearly. In addition, the definition of diminished ovarian reserve (DOR) is not presented and should add up.</p> <p>4. The authors draw the conclusion in the last section: ‘conventional stimulation IVF should be the preferred treatment strategy for most infertile women because it produces the highest live birth rates. Increasingly widely practiced alternative stimulation protocols including minimal stimulation, natural cycle IVF and in vitro maturation (IVM) may have a place in treatment of young women under selective circumstances but appear relatively ineffective in women above age 40 and younger women with DOR’. It is an aggregate data research, simply describing the current usage status of stimulation protocols. This current conclusion needs more evidences to confirm and seems not precise. It is useful to analyze the proportion of component and describe the potential reasons, but the data are limited value to be used as a guide to choose clinical protocols.</p>
--	--

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Reviewer Name: Dr Panagiotis Drakopoulos

Institution and Country: Center for reproductive medicine, UZ Brussel, Belgium Please state any competing interests: No conflict of interest

Please leave your comments for the authors below

I read with great interest the observational retrospective study by Kushnir et al(using the SART database), aiming to evaluate the different protocols used for ovarian stimulation in IVF. It is a hot topic. The authors concluded that conventional stimulation results in significantly higher LBR compared to other protocols (minimal, natural cycle, IVM)

Major remarks

-A severe limitation of the analysis is the absence of standardized definitions for each protocol (i.e. the definition of minimal stimulation is different from the common definition of mild stimulation) and the inability to adjust for relevant confounders.

Therefore, no firm conclusion can be drawn, but only speculations....with a risk of bias

- *Thank you for the comments and careful review. We have modified the conclusion section to be more concise. We agree that there are potential sources of bias and have noted these as important limitations in the discussion section.*

-Another important limitation is that LBR, defined as the 1st fresh or frozen live birth (in freeze all cycles) is not a relevant endpoint in modern ART, if cumulative results cannot be provided.

For each infertile patient, cumulative results (fresh+frozen live birth rates after use of all available embryos) represent the most meaningful outcome. This should be the primary outcome in case we aim to evaluate different stimulation protocols, in the future.

-We agree with this point in principle, however, in practice linking multiple frozen transfers to the initial IVF cycle is very problematic in US registry data. Patients often do not return to use their frozen embryos within a defined time period, they also often move embryos between fertility centers thereby obscuring reliability of cumulative pregnancy rate calculations. Moreover, we point out in the discussion section that "Cumulative live birth rates (per embryo cohort in a single cycle) would, likely, favor conventional stimulation even more profoundly, since these protocols are more likely to result in surplus transferable embryos than any of the alternative protocols."

-Furthermore, the number of cycles (with minimal/natural/IVF) is significantly smaller compared to conventional ovarian stimulation cycles... making the study groups significantly imbalanced. COS cycles account for > 90% of the data. The authors should comment on it, taking also into account the inability to adjust for confounders. As also reported by the authors, the vast majority of minimal/natural IVF protocols were used in 2 US centers, increasing the risk of selection bias.

We have added to the discussion section noting the imbalance in number of cycles between groups. The issue of confounders and bias was addressed in the discussion section already.

We have previously addressed the issue of 2 centers in the manuscript:

Results: "Excluding data from the above mentioned two centers which performed 57.9% of all minimal stimulation and natural IVF cycles showed slightly higher live birth rates (between 0.7% and 6.1%) for these protocols in the remaining centers for all age groups, however, the live birth rates remained significantly lower than those achieved with conventional stimulation."

Discussion: "these two centers reported marginally lower live birth rates than other centers, excluding their data from the analysis did not substantially alter the principal findings."

Minor remarks

-Figure 1: Please provide the number of cycles for each protocol and age strata.

-These data are included in Table 1

I would also suggest to report the results of LBR per protocol and age, in a supplementary Table

-We have added a new Table 2 which provides more detailed information about pregnancy outcomes.

- Further statistical analyses should also be done for comparisons between the other protocols (minimal vs natural etc)

-Below are the p-values for comparison of other protocols to each other to be included as supplemental material. This additional analysis confirms that among young women <35 years of age minimal stimulation produced significantly higher live birth rates than natural cycle or IVM, while there was no difference in live birth rates among other age groups.

	<35	35-37	38-40	41-42	>42
Natural vs. Minimal Stimulation	<0.0001	0.0701	0.2300	0.5798	0.4332
Natural vs. IVM	0.1772	0.9748	0.0944	0.0549	0.4644
Minimal Stimulation vs. IVM	0.0003	0.2869	0.3173	0.0774	0.3707

-Why repeated cycles could not be determined in the analysis, so as to be able to adjust for it? Is this a limitation of the SART database?

-Yes, we only had access to aggregate but not patient level data and therefore could not determine repeat cycles

-Could other causes of infertility be reported?

-Due to the aggregate nature of the data set we were not able to analyze data for other infertility diagnoses

-Please comment on the indication of IVM protocols. Do they include exclusively PCOS patients or also patients with fertilization failures?

-We agree that this is an important point to differentiate indications for IVM protocols. Figure 2 shows live birth rates for patients with DOR diagnosis including those who had IVM. We did not study patients with PCOS diagnosis but agree that this would be an interesting future project and have added a sentence about this to the discussion section.

- Authors should be more conservative with their conclusions in general. The fact the mild stimulation could be an effective option for young good prognosis women in term of fresh live birth rates (not cumulative) is a widely accepted concept (based on meta-analyses). From a clinical point of view, please comment on why COS should be a better choice in women > 42 compared to minimal IVF? I don't think that the analyzed data (with the aforementioned limitations) is robust enough to draw this conclusion

- the words: observational retrospective should be added in the title

- *The title has been modified as suggested.*

-line 56: please add ref 18

-Our version of the manuscript did not have line numbers but ref 18 was added to sentence reading "For that purpose, readers are referred to recent publications" we hope this is the one the reviewer was referring to.

- there is an extra asterisk in Figure 1 for age strata 41-42. To delete please.

-This asterisk was placed to denote that the live birth rate for IVM in this age group was 0%. This data point is also included in the newly added Table 2.

Reviewer: 2

Reviewer Name: Yanping Kuang

Institution and Country: Shanghai Ninth People's Hospital, Shanghai Jiaotong University School of Medicine Please state any competing interests: None declared.

Please leave your comments for the authors below

This is a retrospective cohort study using the aggregate data of U.S. fertility centers for autologous IVF cycles performed during 2014 and 2015 and presented the utilization patterns and live birth rates differ at varying ages with various ovarian stimulation protocols. Although the current data provide valuable information for IVF practice, there are some flaws in the current version:

1. The results presented the limited information including only age, stimulation protocols, whether DOR or not, number of transferred embryos and live birth outcomes which limited the value of this article. More demographic data and cycle characteristics are beneficial to explain the current results. For example, the differences in the definition of stimulation protocols, the characteristics of patients, infertility factors, the treatment or lab techniques among different fertility centers lead to a significant bias. The further stratified analysis and/or the multivariable logistic regression could be useful for live birth rate controlling for these confounding variables. Without these supporting data, the results can not be generalized into the clinical practices, and the result explanations should be with more cautions in the discussion.

-Thank you for the comments and careful review. We recognize that both reviewers have raised the same point and have modified the discussion and conclusion sections accordingly. Unfortunately, the aggregate data do not permit us to control for most confounding variables. Even if we had access to these data points it would be impossible in a retrospective study to control completely for various factors which led to the selection of a particular stimulation protocol by a physician for each patient. It is important to note that by analyzing aggregate national data for the entire population of patients rather than a sample of the population the risk of selection bias is mitigated. Still we agree that there are potential sources of bias and have noted these as important limitations in the discussion section.

2. The live birth rate came from the one ovarian stimulation cycle and fresh embryo transfer, or once stimulation and all subsequent transfer (fresh and frozen embryos) should be described in methods. The difference in live birth outcomes not only come from the stimulation protocols, but also had some confused factors (age, ovarian reserve, gravity, infertility cause etc). The data about the proportion of cancel cycles or no viable embryo transfer are not available.

-We have explained in the methods section that "Live birth rates are now assessed by SART with reference point cycle start, with first embryo transfers considered, whether fresh or the first frozen-thawed transfer in all-freeze cycles"

-We added new data on cycle cancellations and retrievals with no viable embryos to transfer to Table 1.

3. The definition of conventional ovarian stimulation and the minimal ovarian stimulation is not clear. For example, 'the administration of gonadotropins for 8-10 days to recruit multiple mature eggs' and 'use less dose of gonadotropins than conventional ovarian stimulation', the cutoff dose of gonadotropins between conventional ovarian stimulation and minimal stimulation maybe described as a specific range, or provide more detailed data to make it clearly. In addition, the definition of diminished ovarian reserve (DOR) is not presented and should add up.

-We have provided the definitions currently used by SART in national data. The diagnosis and type of stimulation protocols are ultimately entered into the SART system by the treating physician/fertility center. We relied on the definitions used in SART and did not independently attempt to validate these. There are no universally agreed upon cutoff values established for

gonadotropin doses to differentiate conventional and minimal stimulation. Similarly, there are no universally agreed upon definitions for DOR based on age/assay specific ovarian reserve parameters (AMH or FSH levels). We recognize the lack of consistent definitions as a major limitation in the discussion section and the article summary.

4. The authors draw the conclusion in the last section: 'conventional stimulation IVF should be the preferred treatment strategy for most infertile women because it produces the highest live birth rates. Increasingly widely practiced alternative stimulation protocols including minimal stimulation, natural cycle IVF and in vitro maturation (IVM) may have a place in treatment of young women under selective circumstances but appear relatively ineffective in women above age 40 and younger women with DOR'. It is an aggregate data research, simply describing the current usage status of stimulation protocols. This current conclusion needs more evidences to confirm and seems not precise. It is useful to analyze the proportion of component and describe the potential reasons, but the data are limited value to be used as a guide to choose clinical protocols.

-Thank you for the comment, we have revised this section accordingly.

VERSION 2 – REVIEW

REVIEWER	Panagiotis Drakopoulos UZ Brussel Belgium
REVIEW RETURNED	28-Jun-2018

GENERAL COMMENTS	The authors addressed my queries, highlighted the limitations of analyzing registry data and softened their conclusions.
-------------------------	--

REVIEWER	Yanping Kuang Shanghai Ninth People's Hospital, Shanghai Jiaotong University, China
REVIEW RETURNED	25-Jun-2018

GENERAL COMMENTS	This observational retrospective study clearly described that the national utilization patterns and live birth outcomes from various ovarian stimulation protocols in IVF. From the epidemiological piont, the current data had provided some valuable informations for IVF practice. Although the current material had some limitations which did not avoid for causing bias, the authors have given a reasonable explanation.
-------------------------	---