ESHRE Guideline:

Ovarian stimulation for IVF/ICSI

Stakeholder review report

The ESHRE guideline "Ovarian stimulation for IVF/ICSI" was open for stakeholder review between 12 February and 26 March 2019. The draft of the document was published on the ESHRE website. Stakeholders were invited to submit comments through mailings, and advertisements on social media.

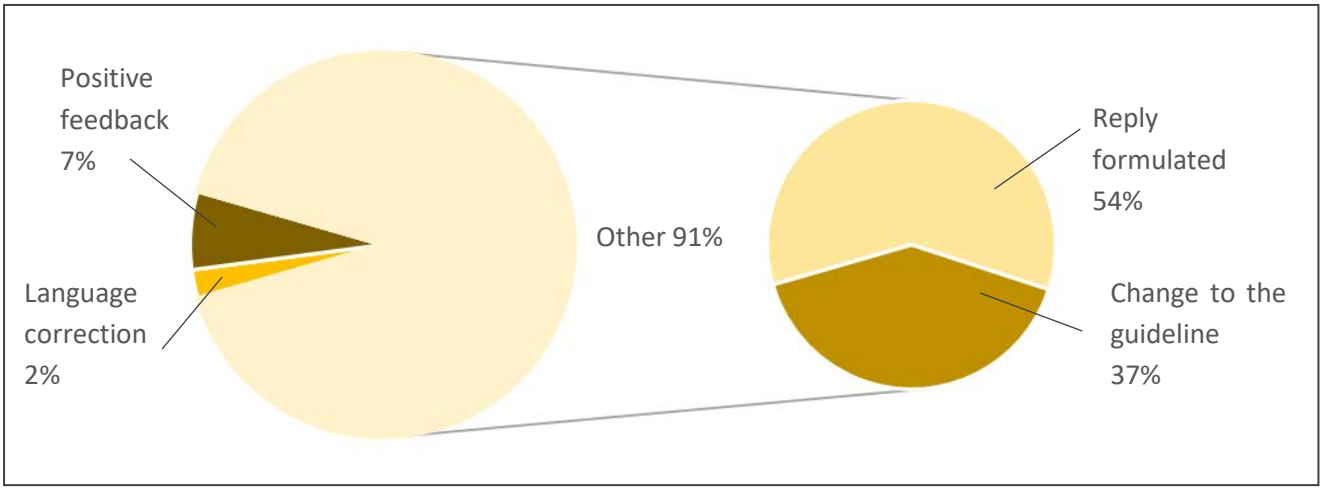
Results

Thirty-nine reviewers, representing twenty countries and 2 national society (The British Fertility Society; EHSRE working groups), submitted a total of 168 comments (on average 4 comments per reviewer). All reviewers are listed on page 2 and in annex 6 of the guideline document.

All comments were assessed by the research specialist and the guideline group members, and, if relevant, changes were made to the guideline (see also Figure 1):

- 4 comments (2.4 %) provided positive feedback that did not require any action from the working group.
- 11 comments (6.5 %) requested improvements of language and format of the guideline, and these were all modified in the guideline
- 153 comments (91%) were comments to the content, requesting corrections, modifications, or addition of further information. Of these, 62 comments were judged relevant and corresponding changes were made to the paper. The working group formulated a reply to the remaining 91 comments, detailing why the comment was not incorporated in the paper.

Figure 1: Results of the stakeholder review: actions for the comments received.



List of reviewers

Representative	Organisations and working groups
Raj Mathur	British Fertility Society
Ferraretti A.P	ESHRE, authors of “ESHRE definition on POR" (2011).
Richard Anderson	ESHRE Fertility preservation guideline group

Reviewer	Country
Hans-Peter Steiner	Austria
Kris Poppe	Belgium
Pratip Chakraborty	India
Mariano Mascarenhas	UK
Arianna D'Angelo	UK
Juan A Garcia-Velasco	Spain
Carlos Calhaz-Jorge	Portugal
Riikka Leppänen	Finland
Ronit Beck Fruchter	Israel
Apostolos Tsironis	UK
Nick Macklon	UK
Klaus Bühler and co-workers	Germany
Hakan Yarali	Turkey
Li Rong	China
Ahmed Samy Saad	Egypt
Aboubakr Mohamed Elnashar	Egypt
Paolo Emanuele Levi-Setti	Italy
Corina Manolea	Romania
Julian Jenkins and co-workers *	Switzerland
Carlo Alviggi and co-workers	Italy and other countries
Thomas D'Hooghe *	Belgium
Filippo Ubaldi and co-workers	Italy and other countries
Fei Gong	China
Yun Lin	China
Jan Olofsson *	Switzerland
Aidong Gong	China
Pablo Diaz-Spindola	Mexico
Ahmet Turp	Turkey
Tamar Barbakadze	Georgia
Sandro C Esteves and co-workers	Brazil
Gemma Castillón Cortés	Spain
William Ledger	Australia
Yuan Li	China
Pedro Barri	Spain
Fabiola Beligotti and co-workers *	Switzerland
Julia Koloda	Russia
Ferraretti A.P.	Italy

Mochtar M., van Wely M., Braat D., Goddijn M.	Netherlands

List of comments

Comments from the industry were also included, however, are indicated with an *

Chapter	Reviewer	Page	Line	Comment	Reply GDG
7	Raj Mathur	19	R 35	This recommendation may be taken as implying that there is a role for adjusting dose before day 6. Perhaps it is better to say that dose adjustment during stimulation is not recommended?	The recommendation was changed to: 'Adjustment (increase or decrease) of the gonadotrophin dose in the mid-stimulation phase during ovarian stimulation is probably not recommended'.
14	Raj Mathur	22	R 61	We feel that this should be clarified with regard to follicle sizes that should be included in the count– is it >18 follicles of all sizes or >18 follicles of >12 mm diameter?	There is no one single diameter of follicles above which we define high ovarian response. For example, in the study by Mathur et al. the size is >12 mm but for Papanicolau and Griesinger ≥ 11 mm, thus we put only the number of follicles without presenting diameter.
17	Raj Mathur	24	/	The GDG should consider whether a recommendation is possible in relation to the value of coasting in agonist and antagonist cycles, as it is a widely practiced method	All interventions in chapter 17 were compared to GnRH agonist triggering of final oocyte maturation. All other interventions were considered outside the scope of the guideline. We will keep your suggestion in mind for the update of the guideline in 4 years.
4B	Raj Mathur	17	R 16	Should this not be a 'strong' recommendation, as it has quite strong evidence backing it?	Most of the studies are old and lack live birth data. We therefore cannot rule out that reduced dosages are at equipoise with the standard dosages of FSH.
4C	Raj Mathur	17	R 21	Given that the evidence shows that a dose of 300 iu gives more oocytes in poor responders than a lower dose, and there is insufficient evidence of an impact on live	In the balance between quality of evidence, the lack of benefit of

				birth rates, it is inconsistent that this is a 'Strong' recommendation. At best this could be a GPP or a 'Conditional' recommendation.	higher dosages than 300 IU and patient preferences and quality of life, the GDG has decided to make this a strong recommendation.
	Raj Mathur	ALL		We would like to congratulate the Guideline Development Group on an excellent, comprehensive and high-quality piece of work.	Thank you.
	Raj Mathur	15 - 25	ALL	The definition of 'Quality of evidence' is not clearly stated. What does one +_ mean relative to four +? There is the potential for confusion when a 'Strong' recommendation may be backed by only 1+ evidence and a 'Conditional' recommendation may be backed by 3+ evidence.	The definitions of the quality of evidence are explained in Annex 2 (summary of findings tables). The definition of a strong and conditional recommendation is explained in annex 5 (methodology).
4C	Hans-Peter Steiner			I am missing my Chapter at Allahbadia Mild IVF	We think this comment refers to Allahbadia G, J Obstet Gynaecol India 2016; 66(5), which is a narrative review, not a systematic review with meta-analysis or comparative study and therefore does not qualify to be added to the body of evidence.
	Hans-Peter Steiner			I am missing a chapter concerning egg collection, which is very heterogeneous worldwide.	Ovarian stimulation is a very broad topic for guideline development. The GDG tried to focus on the most important issues, to avoid making the guideline too extensive. ESHRE has developed a Recommendations for good practice in ultrasound: oocyte retrieval document.
	Kris Poppe			I was wondering why there is no word on the impact of COH on thyroid function in the paper? Thyroid function has an important role to play in reproduction, both before and during pregnancy. A number of studies have shown that COH can lead to an additional strain (on top of that of pregnancy) on the thyroid gland and that following COH, thyroid function can become abnormal ((sublclinical) hypothyroidism); the latter is known to increase the prevalence of pregnancy related complications, including miscarriage. There is indeed no prove that thyroid hormone treatment can improve the efficacy of	Ovarian stimulation is a very broad topic for guideline development. The GDG tried to focus on the most important issues, to avoid making the guideline too extensive. The impact of thyroid function was considered outside the scope of this guideline.

				COH as such, but it can improve a number of pregnancy outcomes in case of hypothyroidism (less MC, less preterm births).	
1	Pratip Chakraborty	15	1	For predicting high and low response to controlled ovarian stimulation, Follicle stimulating hormone and antral follicle count in combination better predicts live birth rate than anti-Mullerian hormone alone in women with diminished ovarian reserve. (Ref: Abstracts of the 31st Annual Meeting of ESHRE, Lisbon, Portugal, 14 June – 17 June, 2015; page: i435)	As stated in the ESHRE manual for guideline development: "the use of abstracts should be avoided except in very rare instances (and always combined with a search for the full paper)"
5	Mariano Mascarenhas	56	26	The recommendation that GnRH antagonist protocol is recommended over GnRH agonist protocol could be subject to a caveat that in women with endometriosis, an ultralong GnRH agonist protocol may provide benefits	The use of GnRH agonist ultralong protocol is based on the increased pregnancy rate observed in a meta-analysis published in 2006 (Sallam et al, Cochrane Database Syst Rev. 2006 Jan 25;(1):CD004635), that includes 3 studies and a total of 165 patients. These 3 studies were published between 1992 and 2002, and none of them used the GnRH antagonist protocol as the control group. Therefore, there is insufficient evidence to recommend its use over the GnRH antagonist protocol in women with endometriosis.
8	Mariano Mascarenhas	73	32	Considering that the Cochrane review suggests an improvement in clinical pregnancy rate with the use of DHEA in poor responder, the grade of the recommendation that DHEA is probably not recommended should probably be downgraded (as there is a possibility that there may be a benefit)	The GRADE mark reflects the quality of the available evidence. Despite the moderate quality evidence, the GDG feels that caution is needed, and well-designed studies are necessary to provide a definite answer.
11	Mariano Mascarenhas	89	14	The COS working group might consider taking into account the evidence that premature progesterone elevation is associated with reduced pregnancy rates after a fresh embryo transfer. Therefore, in selected circumstances, there might continue to be a role for serum progesterone testing on the day of oocyte retrieval.	The association of serum progesterone levels to the achievement of pregnancy has been explored in many studies and meta-analyses and a negative association appears to be present. However, the question examined in the present

					guideline was whether monitoring of ovarian stimulation by ultrasound and progesterone assessment improves safety and efficacy over ultrasound alone. In this respect no recommendation can be made currently in view of the lack of relevant trials.
11	Mariano Mascarenhas	89	14	Additionally, given the evidence that low serum LH levels may be predictive of suboptimal response to GnRH agonist trigger, there might be a role for serum LH testing in women who are planned for GnRH agonist triggering for final oocyte maturation	This is an interesting hypothesis that might need to be explored further. However, in view of the lack of relevant trials examining the value of adding LH to ultrasound for monitoring ovarian stimulation no recommendation can currently be made.
6	Arianna D'Angelo			resp. ??? Please explain abbreviation	We have changed the abbreviation to the full word: 'respectively'
11	Arianna D'Angelo	/	/	There is no recommendation on elevated progesterone serum level at trigger despite there is a huge body of literature and almost all of us test it and freeze if elevated. Is this an oversight?	The association of serum progesterone levels to the achievement of pregnancy has been explored in many studies and meta-analyses and a negative association appears to be present. However, the question examined in the present guideline was whether monitoring of ovarian stimulation by ultrasound and progesterone assessment improves safety and efficacy over ultrasound alone. In this respect no recommendation can be made currently in view of the lack of relevant trials.
12	Arianna D'Angelo	21		Ref 54: I disagree because occasionally there are polyps or hyperplasia which can be seen during the stimulation. The endometrium should always be looked at to exclude potential pathologies.	It was mentioned as a good practice point in the justification that a single ultrasound assessment is necessary

					to identify patients with very thin or very thick EMT, and appropriate diagnostic work-up should be done /example: polyps, hyperplasia, etc./
16	Arianna D'Angelo	/	/	Please specify that this is applicable only to fresh cycles not frozen.Perhaps in the scope of the guideline it should be mentioned that FETs are excluded from this paper.	In the scope of the guideline is stated: "the following issues were outside the scope of the current document: [...], frozen embryo transfer, [...]
17	Arianna D'Angelo	118	3339	there is a recent cochrane review on coasting which should be included	All interventions in chapter 17 were compared to GnRH agonist triggering of final oocyte maturation. The Cochrane meta-analysis on coasting was not included in the body of evidence because it does not compare coasting to GnRH agonist triggering of final oocyte maturation.
4B	Arianna D'Angelo	16		there is no mention to the recent large study from Anja Pinborg which concluded an improved in CPR with agonist but increase in OHSS.	We think this comment refers to the RCT by Toftager et al., 2016; which is excluded from the evidence section of the guideline because is it already included in the meta-analysis by Lambalk et al., 2017. This information is available in annex 7.
4C	Arianna D'Angelo	51		there is a discrepancy between the text and the recommendation. It makes sense the recommendation since low responders should be given 300IU.	The recommendation was adjusted to correct the discrepancy between recommendation and text.
A5	Arianna D'Angelo	129		Page 129 last paragraph spelling mistake" interactive".	This is not a spelling mistake, iterative means that a step-wise process is used when collecting evidence
3	Juan A Garcia-Velasco	38	805	The GDG accepts that estradiol and progesterone are widely used for planning purposes, and considered it acceptable due to the available data on safety and efficacy. But then, it seems a contradiction to contraindicate the use of OCP (strong recommendation). In fact, the literature avoids citing a debate opened by Garcia-Velasco in RBMO, however, the answer to this debate is cited (Griesinger et al). This	It is specified in the introduction that the planning purpose is not addressed in the guideline. All the limitations to the recommendation are explained in the text. The

				contradictory message should be solved, and it would be fair for the potential readers to quote the full debate discussions.	reference to Griesinger was taken out (excluded as it was an opinion review and hence does not qualify to be included for the guideline evidence synthesis).
6	Juan A Garcia-Velasco	63	1580	LH: we are sure that this point was thoroughly discussed among the member of the Study Group, and some RCTs show a benefit in women older than 35 years of age (Bosch et al.). However, this goes unnoticed in the guidelines	According to subanalyses for advanced age in the Cochrane meta-analysis (Mochtar et al., 2017), there is no effect on live birth rate (1 RCT, 240 women) or ongoing pregnancy rate (5 RCT, 1170 women). The study by Bosch et al (2011) was included in that meta-analysis.
9	Juan A Garcia-Velasco	80	2161	Dual stimulation: why should it used only in a research setting? The study group reviewed the available evidence in detail, they agree on random start for fertility preservation in oncological patients and some other indications. When dual stimulation comes into place for low responders, they discuss as a draw back the fact that oocyte or embryo freezing is required. As shown in the studies presented, one of the major indications for dual stim today are low responder patients with advanced maternal age, and a lot of these patients will undergo embryo screening (PGT-A), so yes, freezing is required after embryo biopsy. We would not like dual stim to be in a similar situation of ovarian tissue freezing, which after more than 100 babies have been born, still is considered experimental. A different recommendation should be considered according the available evidence.	We have to look for the best evidence and as mentioned, there is no randomized trial (neither, considered in second hand, retrospective one) comparing dual stimulation with two conventional stimulations in term of the efficacy (cumulative live birth rates or at least number of oocytes) or efficiency (reduced time to live birth). The GDG agrees that in case of urgent fertility preservation (see this section), dual stimulation, when it is possible, is the can be an option to get more eggs in a shorter time. However, with dual stimulation there is a proportion of patient that can get an ongoing pregnancy in the first cycle and don't need the expense of a second cycle.
3	Carlos Calhaz-Jorge	15	R 3-4	Recommendations refer to GnRH antagonist protocol only. Justification states that the evidence applies also to GnRH agonist protocol. Why not to include the latter in the recommendation?	For oestrogen pre-treatment there are no data in GnRH agonist protocol (was adjusted in justification). For progesterone pre-treatment the

					reviewer is correct, and the recommendation was adjusted.
6	Carlos Calhaz-Jorge	18	Last line (following R 33)	A completely different format. Is it intended?	This was indeed intentional, unlike the rest of the table, this is a conclusion and not a recommendation.
6	Carlos Calhaz-Jorge	18	R 27 - 28	Justification refers only to cycle with down-regulation achieved by GnRH agonists. The recommendations seem to apply to any kind of protocols (agonists and antagonists).	The reviewer is correct that the studies in the body of evidence all used the GnRH agonist protocol. Therefore, " in GnRH agonist cycles" was added to the recommendation.
16	Carlos Calhaz-Jorge	23	R 73	In spite of the existence of a meta-analysis (described in page 107, lines 1-4), the recommendation is not clear when to stop progesterone as LPS. Just a vague GPP. Looks odd.	Studies with larger sample sizes are necessary to clearly establish non-inferiority of stopping progesterone supplementation at positive pregnancy test for both GnRH-agonist and GnRH-antagonist protocols.
/	Carlos Calhaz-Jorge	/	/	As a global comment, I'd like to thank the GDG for the terrific job done. To my view this is an excellent document. And well done in graphic organization. That said, I feel that this heavy effort showed clearly the weakness of the available reliable evidence in our field: Quantitatively, out of the 38,840 records identified through database searches only 846 papers were considered eligible for full text assessment. And only 230 of the latter were included in the guideline. Qualitatively, the absence of robust evidence translates in a great number of recommendations stating that "is probably recommended" or "is probably not recommended". Finally, thank you also for identifying several topics for research (Annex III)	Thank you.
/	Carlos Calhaz-Jorge		Title	Controlled ovarian stimulation. The word "controlled" was abandoned in the last version of the International Glossary on Infertility and Fertility Care, and "Ovarian stimulation" was proposed instead. In fact, we'd like to control the process but it's obvious that our capability to do so is very limited.	Adjusted.

				Maybe you can consider to update the title: “Ovarian stimulation in ART” could be an alternative.	
/	Carlos Calhaz-Jorge		Annex 7.	Flowchart 14. Criteria for cycle cancellation: (n=17) is missing in the second dark gray box. It is, out of place, in the second light gray box.	Adjusted.
5	Riikka Leppänen	57	1404	Recommendation “The use of progestin for LH peak suppression is probably not recommended. If applied, progestin can only be used in the context of non-transfer cycles.” I am not familiar with using progestin for LH suppression. I think that clarifying the recommendation may help to understand the way of using progestin or the specific non-transfer cycles you can use the progestin. By doing so it would be possible to comprehend the recommendation without reading the whole text in this chapter.	The use of progestins (MPA:10 mg/day) from stimulation day 1 to the day of trigger has shown to effectively achieve pituitary suppression. The comparative studies have shown similar results to other GnRH analogue protocols. However, the use of this approach involves that a fresh embryo transfer will not be performed (i.e. freeze-all, fertility preservation, oocyte donation)
16	Riikka Leppänen	23	R 77-78	List of all recommendations Chapter 16, 77-78, Justification There is one typographical error: …repeated GnRH agonist infections (correct: injections) alone. The way of writing is correct later in the text p. 112.	Adjusted.
	Riikka Leppänen			I had some difficulties to find the chapter “Abbreviations” (Annex 4). Perhaps this annex would be easier to find from the beginning of the guideline. Or, at least, there could be a reference in the chapter “Terminology”, p. 8, (or “Glossary”, p. 123) to the annex.	We use a standard format for all ESHRE guidelines.
	Riikka Leppänen			I like the way of using italics in chapters to highlight observations from the evidence. I also think that yellow boxes for recommendations are easy to notice.	Thank you.
	Riikka Leppänen	15-24		As time is often limited in a clinical practice, it is important to be able to find the most essential pieces of information easily from the guideline. Therefore, I focused on the list of all recommendations and main topics in the guideline. List of all recommendations I suggest adding more headings to make it easier to find hot topics or adding some kind of colour codes to the number of chapters. It could make it easy to search information from the list of all recommendations. For example, heading “LH suppression and ovarian stimulation” contains 44 recommendations. It would be more convenient to find one specific recommendation if you have a few smaller headings under the main heading.	Adjusted.

10	Beck Fruchter Ronit	84	first sentence	follicular phase twice instead of follicular phase and than luteal phase.	Adjusted.
6	Apostolos Tsironis	64	1631	The comment of Letrozole teratogenicity profile has long been unjustifiable. On the current PCOS guideline Letrozole is recommended as the first line agent for ovulation induction. Therefore, if there are concerns regarding teratogenicity that should be reflected on the PCOS guideline as well. The team's view is that Letrozole is safe and effective both as a sole agent for OI and in combination with gonadotropins mainly for mild approaches in controlled ovarian stimulation. I am aware of large studies currently on the way regarding safety and effectiveness of Letrozole (University of Copenhagen) that show real promise for the use of this agent. I would be grateful if you could review or rephrase your recommendation regarding Letrozole.	There has been a global warning on safety issues regarding this medication, issued by the company which markets it. This warning has not been withdrawn so far. The guideline group chose not to ignore it.
18	Nick Macklon			There appears to be no acknowledgement of the implications that the ongoing move to elective freeze all cycles may have on ovarian stimulation strategies. In such a context, the risk/benefit balance of higher dose stimulation may be altered, and it may be useful for the guideline to address this.	The reviewer addresses the valid point that with Freeze-all as a safety measure, the stimulation in high responders may be driven to levels that still could lead to early OHSS manifestations, especially regarding thrombo-embolism. Currently, research into this topic may be considered absent and a specific question has not been included in the PICO list. We will add this topic for the update of the guideline in 2 years.
	Nick Macklon			The use of the term 'probably recommended' is problematic, as it implies conditionality on some undefined future event or knowledge becoming available, and is therefore difficult to interpret. I would suggest instead using terms such as not recommended, moderately recommended, recommended and highly recommended.	We use universally accepted terminology to formulate recommendations in the ESHRE guidelines, according to GRADE methodology. We will keep your suggestion in mind for the next update of the ESHRE manual of guideline development.

	Nick Macklon			The guideline development group are to be congratulated on this major undertaking.	Thank you.
	Nick Macklon			The text would benefit from correction by a native speaker	We will have the summary paper of the guideline checked by a native speaker before publication
	Nick Macklon			The document would benefit in my view from acknowledging the limitations of basing recommendations solely on research of variable quality and the challenge of extrapolating data from trials to individual clinical situations. This is done from time to time, such as in the context of programming cycles with E2 or P. There is a risk that the approached used is so dismissive of views and practices outwith the north west European perspective that it will not be considered as a balanced view. The position taken on individualizing doing and on limiting doses to 150 IU in expected poor responders are examples of this. There is a need to balance available data with clinicalr ationale.	Thank you for your valuable remark. The GDG attempted to base the recommendations on available evidence whenever possible but acknowledges that this may have been done too strict. The recommendations were reviewed and where possible adapted to be more clinically appropriate, eg the gonadotropin dose recommendation for low responders etc. With regards to the North-west European perspective, experts from south and East Europe where actively recruited, which should have balanced the perspective.
6	Klaus Bühler and co-workers			There are errors in the interpretation of some references, which has led to wrong conclusions (Recommendation: #27)	Based on the comment, it is not clear to the GDG which references are referred to. The GDG has re-examined the recommendation and still stands by it.
6	Klaus Bühler and co-workers			For some recommendations, data from heterogeneous populations were included: young normal responders, aged women, PCOS to reach one general recommendation (Recommendations: #26,28)	For these recommendations, the GDG has looked at the data separately for the general population, PCOS patients and women of advanced age. The conclusions of all studies were similar, which leads to one general recommendation. However, a sentence was added to the justification: 'Studies for this question

					in PCOS and women of advanced age were limited, so that a potential difference between compounds in these subgroups cannot be ruled out based on the current evidence'.
6	Klaus Bühler and co-workers			One recommendation was based on data that were only available in populations treated with GnRH agonists, but this limitation was not made clear in the recommendation. (Recommendation: #27)	The reviewer is correct that the studies in the body of evidence all used the GnRH agonist protocol. Therefore, " in GnRH agonist cycles" was added to the recommendation.
6	Klaus Bühler and co-workers			The use of recombinant FSH (rFSH) and human menopausal gonadotropin (hMG) for controlled ovarian stimulation is equally recommended. • Strength: Strong • Quality of evidence: High to low The recommendation 6.1.1 is based on a Cochrane meta-analysis and several RCTs. We would like to raise our concerns regarding the references that were included, as they referred to studies with a very heterogeneous mixture of study inclusion criteria and study protocol design types: Patient classification, age, treatment protocols and causes of infertility. Devroey et al (2012) included in his study young women receiving ovarian stimulation with a GnRH antagonist protocol. Figen Turkcapar et al (2013) included in their study PCOS patients and used a GnRH Agonist protocol. Ye et al (2012) studied women with advanced age. The meta-analysis of Van Wely (2011) included, among the different study designs, only one RCT with GnRH antagonist protocol, which basically shows results that are opposite to those of Devroey (2012). It is not clear how such mixed evidence-based studies done in different populations can lead to a strong recommendation. This grading seems to be in contrast with the ESHRE Manual for Guideline Development which states that "When the GDG formulates a strong recommendation, they have to be certain about the various factors that influence the strength of a recommendation". Today, with more than 40 years of experience in the ART field, along with the development in technology, and better understanding of the biology of reproduction, we are able to categorize our patients into subgroups and to suggest them a personalized treatment to achieve their one common desire. For many years, the golden standard to measure IVF success rate was the "live birth rate (LBR) per fresh cycle or embryo transfer". The association between the number of retrieved oocytes and LBR, has shown that there is a strong association between the number of oocytes retrieved and live birth rates, at least up to 15 oocytes (Fresh cycles: (Baker et al., 2015;	The guideline group has made a strong recommendation based on a relevant Cochrane review analysing 3197 women. The additional studies retrieved, answering the specific PICO question, were published after the Cochrane review and were not pooled to produce a statistic. These additional studies are in line with the Cochrane conclusion.

				Briggs et al., 2015; Steward et al., 2014; Sunkara et al., 2011). Currently, with the improvement in the freezing and thawing technology, there is increasing consensus (Martins et al, 2018) that we should consider to measure the ART outcome success as "cumulative live birth rates (CLBR)", defined as the first live birth following the use of all fresh and frozen embryos derived from a single ovarian stimulation cycle (Drakopoulos et al., 2016; Malchau et al., 2019; Polyzos et al., 2018), Using this endpoint, it was shown that there is a progressively higher CLBR with increasing number of retrieved oocytes, and the question remains whether a plateau is reached around 15-16 oocytes or more (Malchau et al., 2019) or not (Polyzos et al., 2018; Magnusson et al, 2018). As the number of oocytes retrieved becomes more important when considering CLBR as primary efficiency outcome of ART treatment, it is important to acknowledge that, when compared to hMG treatment in the RCTs considered by the GDG, rFSH treatment results in a higher number of oocytes at egg retrieval (Devroey et al., 2012; Figen Turkcapar et al., 2013) and is associated with a lower gonadotropins consumption (Devroey et al., 2012) and a shorter treatment duration (Figen Turkcapar et al., 2013). A higher number of retrieved oocytes, together with lower gonadotropin consumption and shorter duration of stimulation may lead to reduced costs of ART treatment (Sykes et al., 2001; Fragoulakis et al., 2016). Superiority in oocyte quantity after treatment with rFSH compared to hMG was also shown in many other high-quality studies by (Frydman et al., 2000) (RCT); (Hompes et al., 2008) (An open-label, prospective, randomized comparison of fixed gonadotropin regimens) and Lehert (2010) (Meta-analysis with 16 RCTs, 4,040 patients). And there was also shown a statistically significant increase in live birth rate with GONAL-f compared to u-hMG in patients treated with a long GnRH agonist down-regulation protocol in real world data– an analysis of 24,764 ART cycles in Germany (Bühler 2010).	
6	Klaus Bühler and co-workers	18	rec 28	The use of recombinant FSH (rFSH) and highly purified FSH (hp-FSH) for controlled ovarian stimulation is equally recommended. • Strength: Conditional • Quality of evidence: Low-very low-level evidence The recommendation 6.1.3 is based on a Cochrane meta-analysis (with 13 RCTs (van Wely et al., 2011) and several RCTs (Aboulghar et al., 2010; Gholami et al., 2010; Murber et al., 2011; Parsanezhad et al., 2017; Selman et al., 2010; Selman et al., 2013; Sohrabvand et al., 2012). In the Appendix 2, summary of findings, of these guidelines, the results are reported to be based on low quality evidence because of* (1) High risk of bias associated with poor reporting of methods in one or more primary studies. (2) The pooled effect included the line of no effect and appreciable benefit or harm. (3) Serious inconsistency because	Cumulative live birth rate was chosen as a critical outcome for this guideline. Unfortunately, it is still infrequently reported. Heterogeneity, on the other hand, is something inherent when evaluating clinical trials and for this reason several techniques have been used for its management when synthesizing data. It should be clear though that although it is not

				only 1 Meta-analysis. (4) Small number of events (5) Serious risk of bias due to poor reporting of methodology (see Annex 2, Page 15). It is not clear how such poor-quality evidence could lead to any recommendation at all. Furthermore, the quality of some of the references (Aboulghar et al., 2010; Sohrabvand et al., 2012) used for this recommendation was not evaluated. In addition, on top of the low-quality evidence, we must stress that studies were done in very mixed patient populations, only treated with GnRH agonist protocols. When aiming to develop treatment recommendations for ART patients, it should be made clearer to which patient population the recommendation is applicable, and for which downregulation protocol. The development of rFSH was a big advantage to the fertility landscape as it enables to reduce the inherent variability that resulted from the inconsistent starting materials of urinary FSH production, and to make FSH production independent of urine collection, thus ensuring greater availability. The biological engineering methodology enabled to develop the precise glycosylation pattern of the protein, which provides the highest resemble to the endogenous FSH secreted during the follicular phase of women at their reproductive age. The use of mammalian cell lines with post-translational modification capability resulted in 99% purity recombinant hFSH with homogenous glycosylation pattern, in opposition to the highly purified urinary and pituitary products (Hp-FSH or hMG preparations). The manufacturing process also allows high batch-to-batch consistency with only 2% of variability. This high homogeneity and the follicular phase glycosylation pattern of the recombinant FSH displays higher competency, which results in lower gonadotropins consumption and higher oocyte yield comparing to the urinary products (Andersen et al., 2006; Gonal-F SmPC). As outlined above, the number of oocytes retrieved is positively correlated with the LBR/cycle up to range of about 15 oocytes (Polyzos et al., 2018; Steward et al., 2014; Sunkara et al., 2011). Currently, with the improvement in the freezing and thawing technology, there is increasing consensus (Martins et al, 2018) that we should consider to measure the ART outcome success as “cumulative live birth rates (CLBR)”, defined as the first live birth following the use of all fresh and frozen embryos derived from a single ovarian stimulation cycle (Drakopoulos et al., 2016; Malchau et al., 2019; Polyzos et al., 2018), Using this endpoint, it was shown that there is a progressively higher CLBR with increasing number of retrieved oocytes, and the question remains whether a plateau is reached around 15-16 oocytes or more (Malchau et al., 2019) or not (Polyzos et al., 2018; Magnusson et al, 2018). Based on the elements mentioned above, we are concerned about this	possible to produce a recommendation in the setting of a guideline group based on theoretical arguments, it is possible to arrive at a recommendation in the presence of RCTs. Their quality and their size will define the strength and the uncertainty of a recommendation, as in the current case. The justifications and recommendations here are based on the best possible evidence compiled in reviews and this evidence will still lack the desired high quality, as shown.

				recommendation, as it is based on a heterogeneous patient population and only applies to patients treated with a GnRH agonist protocol during a fresh ART cycle. And also, pharmacoeconomic modelling demonstrated that r-FSH was more cost effective than uFSH when used for ovarian stimulation in the context of ART: having lower total treatment costs vs uFSH, a higher probability of a live birth and a lower cost per live birth (Daya et al., 2001; Fragoulakis et al., 2016; Silverberg et al., 2002).	
11	Klaus Bühler and co-workers		rec 51	The addition of oestradiol measurements to ultrasound monitoring is probably not recommended. • Strength: Conditional • Quality of evidence: Low-level evidence This recommendation is solely based on the Cochrane review by Kwan et al 2014. There is a major concern over the validity of this recommendation to OHSS risk management. The clinically relevant moderate and severe forms of OHSS occur in an estimated 3% to 8% of ART cycles (3% to 6% moderate and 0.5% to 5% severe forms) (Mourad et al., 2017). If we set 5% as the incidence of OHSS, with 5% margin of error and a 95% confidence interval, we will need ~6500 women to generate data that would show difference in OHSS rates (Mourad et al., 2017). However, in the systematic review by Kwan et al only 781 women from 6 studies were included for analysis with ~4% OHSS rate. Therefore, Kwan et al emphasized in their systematic review that "However, these results should be interpreted with caution because the overall quality of the evidence was low. Results were compromised by imprecision and poor reporting of study methodology". No conclusive should be made safely until we have a large well-designed RCT. Yet we should recognize the fact that in many countries/regions oestradiol is routinely checked during ovarian stimulation for minimizing risk of OHSS, as exemplified by the two publications below: Humaidan et al (Humaidan et al., 2016) proposed in his opinion paper in 2016 entitled Ovarian hyperstimulation syndrome: review and new classification criteria for reporting in clinical trials that "elevated or rapidly increasing serum estradiol levels during OS", young age, PCOS, high basal AMH and etc. are "factors associated with an increased risk of OHSS". He also suggested that "treatment in these women should proceed at the lowest effective gonadotrophin dose with routine monitoring (frequent vaginal ultrasonography and/or serum estradiol measurements)." The ASRM 2016 guideline (Practice Committee of the American Society for Reproductive Medicine. Electronic address and Practice Committee of the American Society for Reproductive, 2016) on Prevention and treatment of moderate and severe ovarian hyperstimulation syndrome stated that "serum estradiol concentrations were	A recommendation cannot be based on opinion papers (Humaidan et al., 2016) or other guidelines (ASRM). The question of interest in this case is not whether serum E2 level is associated with OHSS, because this is true. The question of interest is whether the addition of E2 to ultrasound monitoring of ovarian stimulation increases efficacy and safety. In this respect the recommendation made in the current guideline is justified. It is not absolute since the number of patients analysed is not adequate at present and the quality of evidence is low. It is for this reason subject to change in the future if further RCTs show a benefit of E2 assessment in addition to ultrasound evaluation.

				also significantly associated with OHSS…. the mean estradiol value in patients with OHSS was >3,500 pg/mL". This cut-off oestradiol value is associated with increased risk of OHSS (level II-2 evidence) and is Grade B recommendation in the guideline. This is in line with the conclusion made by Kwan et al 2014 (Kwan et al., 2014) that "A combined monitoring protocol including both TVUS and serum estradiol may need to be retained as precautionary good clinical practice and as a confirmatory test in a subset of women to identify those at high risk of OHSS".	
11	Klaus Bühler and co-workers		rec 52	The addition of a hormonal panel consisting of a combination of oestradiol, progesterone and LH measurements to ultrasound monitoring is probably not recommended. • Strength: Conditional • Quality of evidence: Very low-level evidence. This conclusion is based on low-level evidence (two RCTs with 177 women in total) (Golan et al., 1994; Wiser et al., 2012). The sample size of two RCTs are too small (Golan et al:114 and Wiser et al: 63) to make any meaningful comparison with statistical power. In fact, there were only 3 and 4 cases of OHSS from both treatment groups in Golan's study and zero OHSS cases in both treatment groups in Wiser's study. Due to high grade of uncertainty it is hard to make any conclusion on the added value of hormonal monitoring to minimizing OHSS risk. Meanwhile there is a similar trend in both studies showing that more oocytes (Golan et al:13.4 vs 11.7; Wiser et al: 11.7 vs 10) and higher pregnancy rate (Golan et al:25% vs 22.2%; Wiser et al: 57.5% vs 40.0%) were achieved in patients who had transvaginal ultrasound (TVUS) plus hormonal assays during ovarian stimulation as compared to those with TVUS only (not statistically significant different, but clinically important when developing a recommendation). Indeed, the available data from only these two studies are inconclusive. More studies evaluating the optimal procedure for monitoring ovarian stimulation are needed. Meanwhile, monitoring of LH and estradiol levels is critical for ovarian stimulation during a GnRH-antagonist regimen. For example, a serum LH level of 10 IU/L was established as the threshold LH level for GnRH-antagonist to prevent premature LH surge (Borm and Mannaerts, 2000). The flexible use of GnRH-antagonist is also depending on the number and size of growing follicles and/or oestradiol level (Kolibianakis et al., 2011; Lainas et al., 2005). Monitoring of serum progesterone values is also done in standard ART practices because progesterone elevation (PE) on the day of hCG is clearly linked to poor pregnancy outcome after fresh ET in a systematic review (Venetis et al., 2013). More	Due to the drawbacks identified by the guideline group in the published studies the current recommendation is a conditional one and incorporates uncertainty, not however ignoring the fact that currently no extra benefit appears to be present by adding E2 assessment to ultrasound for monitoring ovarian stimulation. There is no doubt that monitoring of LH, oestradiol and progesterone levels provides additional information on follicular growth and endometrium status complementing to ultrasonography as the reviewer suggests. However, this was not the question asked in the current guideline. The question was whether this complementary information improves safety and efficacy. Currently this does not appear to be the case. As clearly stated in the guideline it is not clear if this recommendation is valid for patients treated exclusively with GnRH antagonist due to the lack of relevant studies. The association of serum progesterone levels to the

				recently, Hill et al (2017) reconfirmed this negative correlation in a large retrospective cohort study showing that late follicular phase PE altered ART outcomes in both GnRH agonist and antagonist cycles and reduced pregnancy rate after fresh but not cryopreserved ETs or donor egg ART (Hill et al., 2017). On the other hand, a low serum progesterone level is also a sign of poor follicular growth during ovarian stimulation. A single center retrospective cohort study showed that progesterone levels ≤0.5 ng/ml on the day of hCG administration was associated with lower live birth rates (Santos-Ribeiro et al., 2014). In consistency with this finding, a Belgian single center RCT showed that delaying ovulation triggering by 24 h for patients with 3 leading follicles ≥18 mm but low progesterone levels (<1 ng/ml) lead to more mature oocytes retrieved (mean difference 2.41, P=0.031) (Vandekerckhove et al., 2014). Thus, monitoring of progesterone level (at least) on the day of triggering may provide additional information on patient prognosis and facilitate personalized ovulation triggering and ET strategy. To conclude, monitoring of LH, oestradiol and progesterone levels provides additional information on follicular growth and endometrium status complementing to ultrasonography. We propose that this recommendation can be removed until better evidence from large well-designed RCTs is available, as the absence of greater efficiency, based on poor quality studies, should not be translated in a recommendation against a specific method of monitoring.	achievement of pregnancy has been explored in many studies and meta-analyses and a negative association appears to be present. However, the question examined in the present guideline was whether monitoring of ovarian stimulation by ultrasound and progesterone assessment improves safety and efficacy over ultrasound alone. In this respect no recommendation can be made currently in view of the lack of relevant trials.
13	Klaus Bühler and co-workers			Some recommendations have been made without any supportive evidence provided (Recommendations: #57,58)	These recommendations were converted to GPP.
13	Klaus Bühler and co-workers		rec 57	It is not recommended to base timing of final oocyte maturation triggering on oestradiol levels. • Strength: Strong • Quality of evidence: Very low-level evidence.	These recommendations were converted to GPP.
13	Klaus Bühler and co-workers		ec 58	It is not recommended to base timing of final oocyte maturation on oestradiol/follicle ratio. • Strength: Strong • Quality of evidence: Very low-level evidence. Comments: There are no interventional studies investigating triggering based on the oestradiol levels or oestradiol/follicle ratio. But as mentioned in our comments to recommendation No. 51, the serum estradiol level is critical for adjusting ovulation triggering strategy to minimize OHSS risk. This is also in line with recommendation No.	These recommendations were converted to GPP.

				56 that “the decision on timing of triggering in relation to follicle size is multi-factorial, taking into account the size of the growing follicle cohort, the hormonal data on the day of pursued trigger……”. Again, the absence of greater efficacy should not be translated in a recommendation against a specific method of monitoring. We therefore propose that recommendations No. 57 and 58 are removed until we have solid data from large well-designed RCT showing no added value of estradiol levels for the timing of ovulation triggering.	
4C	Klaus Bühler and co-workers			A higher gonadotropin dose of 300 IU is probably not recommended over the conventional dose of 150 IU for predicted low responders. • Strength: Strong • Quality of evidence: low The recommendation 4.1.C is based on a Cochrane meta-analysis with 2 RCTs; (Lensen et al., 2018). In the Appendix 2, summary of findings, of these guidelines, the results are reported to be based on low quality evidence because of: 1. Serious risk of bias due to incomplete reporting of methodology in individual studies 2. (2) The pooled effect included both the line of no effect and appreciable benefit or harm (see Annex 2, Page 10). It is not clear how such poor-quality evidence could lead to a strong recommendation. This grading seems to be in contrast with the ESHRE Manual for Guideline Development which states that “When the GDG formulates a strong recommendation, they have to be certain about the various factors that influence the strength of a recommendation”. Furthermore, the meta-analysis by Lensen et al., 2018, addressed a wide range of dose comparisons, but in fact there were only 2 trials that tested the comparison between 150 IU and 300/450 IU gonadotropins. 1. Klinkert ER (low quality RCT, with only 52 participants): The aim of the study was to evaluate the effect of doubling the starting dose of gonadotrophins on the ovarian response in IVF patients with a low antral follicle count (AFC), namely 150 IU compared to 300 IU. For that, they have randomized 52 patients with an AFC of <5 follicles to 2 groups: 150 IU (group I, n = 26) or 300 IU (group II, n = 26) of rFSH as a starting dose. The main outcome measures of the study were number of oocytes, poor response and ongoing pregnancy. The conclusion was that expected poor response patients, defined as patients with an AFC <5, are likely not to benefit from a higher starting dose of gonadotrophins in IVF. However, it must be noted that 9 patients (34%) in the group of the 150 IU had dose adjustments due to lack of ovarian response, whereas no dose	The recommendation was changed to ‘it is unclear whether a higher gonadotropin dose is recommended over 150 IU. Despite the higher number of oocytes, there was no difference in live birth/ongoing pregnancy rate. Furthermore, the sample sizes of the studies are small and therefore not sufficient to provide evidence for dose comparisons for live birth outcome.

				adjustments were needed in the group of the 300 IU (Klinkert et al., 2005). 2. Van Tilborg TC: OPTIMIST, examined also whether an increased FSH dose resulted in higher cumulative live birth rates in women with a predicted poor ovarian response, with low antral follicle count (AFC), scheduled for IVF or ICSI. For that they have randomized 511 patients as follow: Women with an AFC ≤ 7 were randomized to an FSH dose of 450 IU/day or 150 IU/day, and women with an AFC 8-10 were randomized to 225 IU or 150 IU/day. They concluded that for women with a predicted poor ovarian response (AFC < 11) undergoing IVF/ICSI, an increased FSH dose (225/450 IU/day) does not improve cumulative live birth rates as compared to a standard dose (150 IU/day; van Tilborg et al., 2017). Looking in depth into these 2 only studies that were considered for this recommendation, several concerns can be raised: 1) Both studies were performed in patient populations (AFC<5, AFC <=7) that did not meet the definition of Low ovarian response used by the GDG "Low ovarian response is a diminished response to conventional ovarian stimulation, characterized by the presence of a low number of follicles and/or oocytes (Ferraretti et al., 2011). Generally, ≤ 3 follicles on day of oocyte maturation trigger and/or ≤ 3 oocytes obtained characterize a low response" 2) Cycle cancellation is an important risk factor that should be considered from an effectiveness and patient point of view, as one of the major concerns for low ovarian responders is treatment cancellation due to lack of response. Using doses of 150 IU for this population resulted in 34% dose adjustment in the study reported by Klinkert et al (2005), and in ~30% cycle cancellation in the RCT performed by Van Tilborg et al (2017), where dose adjustment was not allowed. The OPTIMIST trail by Van Tilborg (2017), was widely criticized by many leading experts from the global ART community with respect to: • patient population studied: the definition of POR does not fulfil neither the established ESHRE Bologna POR criteria (Ferraretti et al., 2011) nor the recently suggested POSEIDON criteria for predicted POR (Poseidon et al., 2016) [see:. Haahr et al (2018)]. • cancellation rates and Number of oocytes: It is fact that without individualization there is a significantly higher cancellation rate and lower retrieved oocytes [see: Sunkara & Polyzos (2018); La Marca (2018)]. It is well known "…that individualization of ovarian stimulation reduces the variability of the	

				number of oocytes recovered, increases the number of oocytes recovered in the poor responder ….and reduces the risk of cancellation of the cycle" (La Marca et al., 2018) To conclude, both RCTs that are the base of this recommendation, suggests that for predicted low responders, a higher dose of gonadotropins results in more oocytes and reduces the risk of cycle cancellation.	
	Klaus Bühler and co-workers			Conclusion regarding number of oocytes was based on references referring to completely different patient populations (IUI and IVF), which cannot be pooled regarding this intermediate treatment outcome, namely No. of oocytes (page 12). The discussion regarding the number of oocytes is essentially an ART (for IVF and ICSI treatment) discussion, as it is well accepted that, in the context of IUI, ovarian stimulation should not aim to stimulated more than 1 follicle or 2 follicles	There were indeed 2 references listed referring to IUI patient population. This was a mistake and has been corrected.
	Klaus Bühler and co-workers			Although cumulative live birth rate (CLBR) was stressed by the GDG as a critical outcome for these guidelines, the recommendations did NOT refer to publications reporting CLBR; for that reason, the recommendations need to specify that they apply to fresh cycles only.	For very few recommendations CLBR was reported. The recommendations are always based on the critical outcomes, as formulated in the scope section of the guideline. Available evidence regarding the critical outcomes is described in the evidence section of each recommendation.
	Klaus Bühler and co-workers	pag 12		The authors cannot reach this conclusion, as they used evidence from mixed treatments in Medical Assisted Reproduction (IUI and ART), and it is well known that the number of oocytes should be limited in the context of IUI and should be optimized in the context of IVF. Therefore, it is not acceptable that they used both references related to IUI patients (Cantineau et al., 2007; Ragni et al., 2004) and ART patients (Lensen et al., 2018; Sterrenburg et al., 2011) as a scientific basis. Moreover, the authors failed to consider and consider the many publications reporting a positive association between the number of oocytes and Cumulative Live birth Rate (CLBR) and or Live Birth Rate (LBR) after ART treatment. Relevant publications reporting a positive association between number of oocytes and live birth rates, Cumulative live birth rate after ART treatment. a) Oocytes No. and LBR: A positive association between the number of retrieved oocytes and LBR has been reported in many studies. Recently, Toftager (2017)	The two references were used to underline the information on the relation between FSH dosage and follicle and oocyte number. This relation has been studied also outside of the IVF/ICSI context. However, as it may cause confusion, these references were taken out of the introduction. Regarding the correlation between oocyte number and live birth rates, we need to make clear that these cross-sectional correlation data may not automatically imply a causal relationship between oocyte number

				performed a rigorous RCT with >1000 participants and reported a correlation between the number of oocytes retrieved and an increased chance of live birth. In addition, many other retrospective studies, based on big data, confirmed the strong association between the number of oocytes retrieved and live birth rates, at least up to 15 oocytes (Fresh cycles:(Briggs et al., 2015; Steward et al., 2014; Sunkara et al., 2011)). b) Oocytes No. and CLBR: Three excellent studies have reported a positive association between the number of retrieved oocytes and CLBR, defined as the first live birth following the use of all fresh and frozen embryos derived from a single ovarian stimulation cycle (Drakopoulos et al., 2016; Malchau et al., 2019; Polyzos et al., 2018), Using this endpoint, it was shown that there is a progressively increasing CLBR with increasing number of retrieved oocytes, and the question remains whether a plateau is reached around 15-16 oocytes or more (Drakopoulos et al., 2016; Malchau et al., 2019; Polyzos et al., 2018) or not (Polyzos et al., 2018). c) Oocytes No. and embryo quality: In a recently published systematic review presented at the ASRM meeting in 2018, it has been clearly demonstrated that there is a positive association between the number of eggs obtained at egg aspiration, and the number of good quality embryos (Day 3, Day 5, euploid embryos) (D'Hooghe et al, 2018) As of the stated above we would suggest changing this statement to: "In the context of ART treatment, there is a positive correlation between number of eggs, LBR per fresh cycle and CLBR per started cycle".	and live birth rate. The only way to show this is with RCTs, comparing distinct dosage and thereby oocyte number levels. If in this comparison more oocytes in the higher dosage trial arm create more babies, then we have proven that the correlation is indeed without confounding. As of today, none of such studies has ever supported this. Interestingly, the Toftager trial demonstrated that with lower oocyte number, by using the antagonist system, there were not more live births. We have discussed this issue more extensively now in the general introduction.
	Klaus Bühler and co-workers			In these proposal for ESHRE guidelines for "Controlled Ovarian Stimulation for IVF/ICSI" 87 statements are included. In 34 cases (39%) the conclusion is: "probably not recommended" (27; 31%) or "probably recommended" (7; 8%). And looking only to the chapter "LH suppression and ovarian stimulation" we see all in all 43 statements with 19 cases (44%) of "probably" (not) recommended and 1 case of insufficient evidence. Where can be seen the value of such guidelines? The Cambridge Advanced Learner's Dictionary & Thesaurus defines guidelines as: "information intended to advise people on how something should be done or what something should be." But where is the determination if in nearly half of the statements it is not said in which	Thank you for your valuable remark. "Conditional" recommendations (fi "probably recommended") are well accepted in international guideline development methods and should be applied to situation where there is a need to critically assess the benefits and harms for the individual patient. We agree with the limitations mentioned, ideally, each of the recommendations would be a strong recommendation based on high

				direction one should go? Probably left? Probably right?, or probably in a complete other direction? We think that such statements are very embarrassing for an institution like ESHRE. Such statements do not fulfil their quality requirements and do not suit their levels. With nearly half of the recommendation classified as "conditional", and many of all the recommendations are based on (Cochrane) meta-analysis of which the authors are specially reporting POOR or low evidence, this guidelines are pre-mature and should be viewed as studies to be done. So, they do not constitute a guide or aid in daily work in reproductive medicine. We doubt that the guidelines in this form will do justice to EHSR's worldwide excellent reputation and live up to the own ambitious standards.	quality evidence, but unfortunately this evidence was not available for the majority of recommendations. Within these limitations, the guideline does provide valuable advice on how to approach ovarian stimulation, while at the same time highlighting the areas of uncertainty. This should stimulate researchers to fill the research gaps and allow more strong and clear recommendations when the guideline is updated.
	Klaus Bühler and co-workers			The proposed strength of several recommendations (Recommendations: #20,27,28,51,52,57,58) does not correspond to the quality of the evidence presented	The strength of recommendations is based on a framework, taking into account the evidence, balance between favourable and unfavourable effects and acceptability to stakeholders and patients. The strength of recommendations #20, #57 and #58 were adjusted. The other recommendations were re-evaluated, however, the GDG still stands by these recommendations.
9	Hakan Yarali	80-81	2145-2172	We recognize that there are no prospective randomized trials (RCT) that compare dual stimulation with two conventional stimulations in terms of efficacy (cumulative live birth rates) or efficiency (reduced time to live birth) of the two strategies. We also recognize that mandatory freeze-all of oocytes or embryos may be a disadvantage of this protocol because of additional procedure and oocyte manipulation, which, may not be allowed by some national health care. Nevertheless, it must be noted that freeze-all is mandatory also in case of luteal phase stimulation-only, random start, oocyte/embryo accumulation through sequential conventional stimulations and blastocyst stage PGT-A cycles. In addition, we do not understand why Committee Members did not mention in the evidence section that, according to all the papers published on the topic, the mean	We have to look for the best evidence and as mentioned, there is no randomized trial (neither, considered in second hand, retrospective one) comparing dual stimulation with two conventional stimulations in term of the efficacy (cumulative live birth rates or at least number of oocytes) or efficiency (reduced time to live birth). The GDG agrees that in case of urgent fertility preservation (see this section), the

				number of oocytes retrieved in the luteal phase stimulation is significantly higher than follicular phase as are the mean number of blastocysts and of euploid blastocysts. Moreover, the chance to find an euploid embryo or a blastocyst to transfer is significantly higher per started ovarian cycle in the dual stimulation if compared to standard stimulation (Ubaldi et al. 2016). Finally, dual stimulation is applied successfully by many centers in different countries. And the evidence published in favor of this procedure in increasing day by day (Xu and Li, 2013; Kuang et al., 2014; Moffat et al., 2014; Ubaldi et al, 2016; Wei Li-Hong et al., 2016; Tsampras et al., 2017; Vaiarelli et al, 2017; Cardoso et al., 2017; Liu et al., 2017; Cimadomo et al., 2017; Zhang Wei et al., 2018; Rashtian and Zhang 2018; Madani et al., 2018; Bailing Jin et al 2018; Vaiarelli et al 2018; Sighinolfi, Sunkara, La Marca, 2018; Alsbjerg, Humaidan et al., 2019). While we understand that this procedure cannot be suggested for standard patients, poor prognosis patients (e.g., with reduced ovarian reserve, AMA, Bologna POR) might benefit from it. In conclusion, although there are no RCTs that show the superiority of dual stimulation vs conventional stimulation in terms of efficacy and efficiency, the author of this guideline could not undervalued the available evidence we believe that there are enough clinical data to state that "dual stimulation can be considered in poor prognosis patients when freeze-all is mandatory. It is not clear why the Committee stated that: "Luteal phase stimulation could be used in the non transfer cycles" although it has far less clinical and laboratory evidence (some of which use data from dual stimulation) reported in the literature.	dual stimulation, when it is possible, is the can be an option to get more eggs in a shorter time. However, in the dual stimulation there is a proportion of patient that can get an ongoing pregnancy in the first cycle and don't need the expense of a second cycle.
1	Li Rong	31	578	Recommendation For predicting high and low response to controlled ovarian stimulation, use of either antral follicle count (AFC) or anti-Müllerian hormone (AMH) is recommended over other ovarian reserve tests. Comment: A recent prospective study shows a significant longitudinal fluctuation in AMH levels per participant, expressed as coefficient of variation (CV) intra-cycle of 20.7%, and a 28% of variation between the AMH values measured on day 2/3 of two consecutive menstruations [Front Endocrinol (Lausanne). 2018 Nov 27;9:686]. Due to such significant fluctuation and variation of AMH, treatment decision can not be easily made if AMH is used alone for predicting ovarian response.	This recommendation is based on the results of an IPD meta-analysis in which is demonstrated that a single AFC or AMH measurement shows a high accuracy in the prediction of response. Despite of any fluctuations that may exist. The accuracy of AFC and AMH are higher compared to other ORTs. As stated in the justification we did not compare the tests or studied the added effect of using multiple tests.
16	Ahmed Samy Saad	108	3	In a prospective and randomized trial, Aboulghar et al. (2008) demonstrated that luteal phase support can be interrupted with no complications after the first positive ultrasound at 6-7 weeks of pregnancy.	The study that is mentioned here is included in the meta-analysis by Liu

				So, we may consider continuing whatever the route of luteal phase support progesterone used, till the first positive ultrasound scan. This is assuring for the patient & we don't risk to interrupt the luteal phase support with the risk of miscarriage Aboulghar MA, Amin YM, Al-Inany HG et al. Prospective randomized study comparing luteal phase support for ICSI patients up to the first ultrasound compared with an additional three weeks. Hum Reprod 2008; 23: 857–62.	et al. 2012, which is included in the body of evidence.
16	Ahmed Samy Saad	108	5	From the studies of the drugs which are the base of the pharmaceutical companies recommendations on the use of the drugs,the recommendations of its use from the manufacturer is different from the broad recommendation to use the progesterone till the pregnancy test only, for example: 1- The dose of ENDOMETRIN is 100 mg administered vaginally two or three times daily starting the day after oocyte retrieval and continuing for up to 10 weeks total duration. 2- Your doctor may prescribe one applicator of CRINONE either daily or twice daily and will advise you when to start treatment. If pregnancy occurs treatment may continue for up to 12 weeks. 3- Prolutex used for 10 wks duration and for 7 wks but not for preg test	The GPP was adjusted and should now mention all available formulations.
7	Aboubakr Mohamed Elnashar	68	1770	Increase of the gonadotrophin dose beyond stimulation day 6 during controlled ovarian stimulation is probably not recommended. Because we can decrease dose to prevent OHSS. Please delete decrease.	To decrease the risk of OHSS a lower FSH dose is recommended from the start of stimulation in GnRH agonist cycles (cfr. Rec. 4A.12). We have no data regarding decreasing FSH doses after day 6 of ovarian stimulation in relation to OHSS risk.
12	Aboubakr Mohamed Elnashar	91	2531	Endometrial pattern is more sensitive than endometrial thickness in reflecting endometrial receptivity	In the literature there are studies summarizing more parameters of endometrium /uterine ultrasonographic scoring system/: thickness, pattern, power doppler, contraction, etc. Firstly different endometrial type classifications have been used, so the conclusions would be inaccurate, and secondly, the key question and the scope of guideline is only endometrial thickness during COS.

12	Aboubakr Mohamed Elnashar	91	2532	Counseling on day of triggering not day of oocyte pick up (too late). Plese delete on day of pick up	Some studies included in evidence table even make the EMT assessment also on the day of the embryo transfer, which automatically means that the assessment on the pick-up day is not late.
13	Aboubakr Mohamed Elnashar	93	2598	Trigger at size 16-22 is for agonist protocol not antagonist, antagonist 15 to 17 mm	The guideline reads: Most often, final oocyte maturation is triggered at sizes of several of the leading follicles between 16-22 mm. This recommendation is compatible with both GnRH-agonist and GnRH-antagonist based ovarian stimulation. A more specific recommendation cannot be inferred from the existing literature.
14	Aboubakr Mohamed Elnashar	97	2700	The diameter of the follicles should be mentioned	This was adjusted to:,In a large prospective cohort study with 1801 women (2524 cycles), the threshold of≥18 follicles ≥ 11 mm…
4A	Aboubakr Mohamed Elnashar	45	992	A reduced gonadotropin dose is recommended to decrease the risk of OHSS in predicted high responders if GnRH agonist protocols are used. Oudshoorn, et al., 2017 study was done on antagonist and agonist protocols, So please add or antagonist protocols	The justification states that the recommendation is extrapolated from a stratified group analysis of the RCT in which majority of the patients were treated with the long GnRH agonist protocol.
4C	Aboubakr Mohamed Elnashar	51	1180	A higher gonadotropin dose of 300 IU is probably not recommended over the conventional dose of 150 IU for predicted low responders. This is to be deleted because it counteracts the next strong recommendation. A gonadotropin dose higher than 300 IU is not recommended for predicted low responders. So the maximum dose is 300 IU	The recommendation was changed to 'it is unclear whether a higher gonadotropin dose is recommended over 150 IU. Despite the higher number of oocytes, there was no difference in live birth/ongoing pregnancy rate. Furthermore, the sample sizes of the studies are small and therefore not sufficient to

					provide evidence for dose comparisons for live birth outcome.
9	Paolo Emanuele Levi-Setti	80-81	2145-2172	We recognize that there are no prospective randomized trials (RCT) that compare dual stimulation with two conventional stimulations in terms of efficacy (cumulative live birth rates) or efficiency (reduced time to live birth) of the two strategies. We also recognize that mandatory freeze-all of oocytes or embryos may be a disadvantage of this protocol because of additional procedure and oocyte manipulation, which, may not be allowed by some national health care. Nevertheless, it must be noted that freeze-all is mandatory also in case of luteal phase stimulation-only, random start, oocyte/embryo accumulation through sequential conventional stimulations and blastocyst stage PGT-A cycles. In addition, we do not understand why Committee Members did not mention in the evidence section that, according to all the papers published on the topic, the mean number of oocytes retrieved in the luteal phase stimulation is significantly higher than follicular phase as are the mean number of blastocysts and of euploid blastocysts. Moreover, the chance to find an euploid embryo or a blastocyst to transfer is significantly higher per started ovarian cycle in the dual stimulation if compared to standard stimulation (Ubaldi et al. 2016). Finally, dual stimulation is applied successfully by many centers in different countries. And the evidence published in favor of this procedure in increasing day by day (Xu and Li, 2013; Kuang et al., 2014; Moffat et al., 2014; Ubaldi et al, 2016; Wei Li-Hong et al., 2016; Tsampras et al., 2017; Vaiarelli et al, 2017; Cardoso et al., 2017; Liu et al., 2017; Cimadomo et al., 2017; Zhang Wei et al., 2018; Rashtian and Zhang 2018; Madani et al., 2018; Bailing Jin et al 2018; Vaiarelli et al 2018; Sighinolfi, Sunkara, La Marca, 2018; Alsbjerg, Humaidan et al., 2019).	We have to look for the best evidence and as mentioned, there is no randomized trial (neither, considered in second hand, retrospective one) comparing dual stimulation with two conventional stimulations in term of the efficacy (cumulative live birth rates or at least number of oocytes) or efficiency (reduced time to live birth). The GDG agrees that in case of urgent fertility preservation (see this section), the dual stimulation, when it is possible, is the can be an option to get more eggs in a shorter time. However, in the dual stimulation there is a proportion of patient that can get an ongoing pregnancy in the first cycle and don't need the expense of a second cycle.
	Paolo Emanuele Levi-Setti	9	241-256	GDG refer to 3 categories of response low (less than 4 oocytes), normal (4-19 oocytes), high (more than 19 oocytes). The proposed guidelines forget completely data and meta-analysis on intermediate (4-8 oocytes) and the research (big-data, retrospective and prospective trials) on different affects of agonist-antagonist protocols, starting dose and LH+ FSH effect in this category of patients, representing more than 40% of our infertile population enrolled in ART cycles.	The GDG is very aware that there are more nuanced definitions of a low ovarian response but decided to try and keep it simple: in the guideline there is only low-normal-high responder. Furthermore, only few studies have been performed in this intermediate group. Therefore, the

					GDG only suggested definitions for low and high response.
	Paolo Emanuele Levi-Setti	11-12	312-359	When considering the number of oocytes or better the number of mature/MII oocytes a more recent evaluation of the frozen blastocysts delivery rate must be fully considered and how CDR is influenced in poor responders even by one more oocyte and in high responders even by more than 25 oocytes retrieved. A full consideration of the couples' procreative plan when more than 2 children (even more than 5 in our experience) can be obtained by a single retrieval during a period as long as more than 10 years, behind every chance of a positive new cycle. Vaughan et al. Fertil Steril_ 2017;107:397–404.	Regarding the correlation between oocyte number and live birth rates, we need to make clear that these cross-sectional correlation data may not automatically imply a causal relationship between oocyte number and live birth rate. The only way to show this is with RCTs, comparing distinct dosage and thereby oocyte number levels. If in this comparison more oocytes in the higher dosage trial arm create more babies, then we have proven that the correlation is indeed without confounding. As of today, none of such studies has ever supported this. Interestingly, the Toftager trial demonstrated that with lower oocyte number, by using the antagonist system, there were not more live births... We have discussed this issue more extensively now in the general introduction.
9	Corina Manolea	80	2161	The recommendation of restricting double ovarian stimulation to clinical research seems too tight in today`s practice when clinicians are confronted with a lot of poor responders, often of advanced age, often with previous failures of conventional COS, and for whom very little research has been undertaken since the first description of the POR in the early 80`S. Also, it does not seem to be in line with previous recommendations on LP stimulation or random start, which, in fertility patients, are not restricted to clinical research according to this guideline. After proper counseling, double ovarian stimulation could be offered as an alternative treatment to poor responders of advanced reproductive age that yield very low numbers of eggs (0--3) during the follicular phase and have 1 or more previously failed IVF cycles with early follicular phase stimulation. Until cost-effectiveness studies	We have to look for the best evidence and as mentioned, there is no randomized trial (neither, considered in second hand, retrospective one) comparing dual stimulation with two conventional stimulations in term of the efficacy (cumulative live birth rates or at least number of oocytes) or efficiency (reduced time to live birth). The GDG agrees that in case of urgent fertility preservation (see this section), dual

				emerge, double stimulation should be offered as a “rescue” treatment to selected cases of poor responders; in such cases, it could increase the possibility of obtaining at least one euploid embryo (Ubaldi et al, 2016) and reduce drop out rates that are highest in these patients. Feasability of COS in a high endogenous or exogenous P4 environment is supported by research of the past 10 years; neonatal outcomes seem unaffected (Wang et al 2018) and the rapid widespread of non-conventional ovarian stimulation in real life clinical practice is showing that clinicians worldwide see advantages of these approaches (at least) for some of their low responder patients. The drawback of mandatory freeze-all should be judged in the context of today´s routine use of vitrification.	stimulation, when it is possible, is the can be an option to get more eggs in a shorter time. However, with dual stimulation there is a proportion of patient that can get an ongoing pregnancy in the first cycle and don’t need the expense of a second cycle.
	Corina Manolea	127	3515	Clinical research on non-conventional ovarian stimulation: -cost-effectiveness of double ovarian stimulation in the same cycle as compared to 2 IVF cycles of conventional stimulation in poor ovarian responders (time-to-pregnancy should be also considered). - optimization of ovarian stimulation during the luteal phase as regards the moment of initiation of gonadotrophins, the starting dose, timing of antagonist administration i(n order to reduce FSH consumption and stimulation period).	Thank you for the suggestions.
6	Julian Jenkins *	60	3	To clarify there are 3 distinct rFSH options it would be helpful at the start of this section to add the text below 6.1 RECOMBINANT FSH (RFSH) rFSH is available as follitropin alpha, originator and biosimilar products, and follitropin beta that all may be used in a similar manner guided by a cumulative large body of evidence on clinical use. In addition, there is a long acting rFSH, corifollitropin alfa, and a further rFSH, follitropin delta, which has an individualised daily dosing regimen based on a woman’s serum anti-Müllerian hormone level and her body weight.	The tools available for increasing FSH exposure are several, but basically most comprise preparations containing FSH. The source can be urinary (purified or highly purified) or recombinant. Some preparations combine FSH with of LH, or LH like activity. The vast majority of FSH compounds are distributed for to dosing in International Units, as standardisation based on an oestradiol output bio-assay. Only one compound is delivered in micrograms, and dosing here is based on a dosing algorithm. We have added this point to the guideline text in the introduction.

16	Julian Jenkins *	23	3	Re: The dosing of natural progesterone has evolved empirically, usually dosages used include ... Please note: Following a full development programme including a dose finding phase 2 study and a phase 3 study through a decentralised procedure in January 2017 all EU member states agreed to grant market authorisations for Cyclogest® for luteal phase support as part of assisted reproduction treatment at a dose of Cyclogest® 400 mg pessary administered intravaginally twice daily http://www.mhra.gov.uk/home/groups/par/documents/websiteresources/con775404.pdf Accordingly, we would recommend adding to the dosing recommendation for natural progesterone: 400 mg progesterone vaginal pessary twice daily	The GPP was adjusted and should now mention all available formulations.
16	Julian Jenkins *	106	8	Dosing Duijkers 2018 study had not been published at the time of the van der Linden 2015 Cochrane review and as this study has critical information of relevance for progesterone dosing it should be considered in this section. Specifically, this phase 2 dose finding study showed that the best secretory transformation of the endometrium was observed during treatment with 400 mg progesterone vaginal pessaries, administered twice daily. Once daily dosing with progesterone vaginal pessaries were found to be inferior to twice daily dosing irrespective of the dose. Duijkers IJM, Klingmann I, Prinz R, Wargenau M, Hrafnsdottir S, Magnusdottir TB, Klipping C Hum Reprod. Effect on endometrial histology and pharmacokinetics of different dose regimens of progesterone vaginal pessaries, in comparison with progesterone vaginal gel and placebo. 2018 Nov 1;33(11):2131-2140. doi: 10.1093/humrep/dey288.	The GPP was adjusted and should now mention all available formulations.
16	Julian Jenkins *	107	24	Re: Recommendations The dosing of natural progesterone As explained above there is now further evidence to guide ART practice beyond the van der Linden 2015 Cochrane review. Accordingly, we would recommend adding to the dosing recommendation for natural progesterone: 400 mg progesterone vaginal pessary twice daily	The GPP was adjusted and should now mention all available formulations.
6	Carlo Alviggi	62-63	1553-1588	The strength of the recommendation for low responders and women of advanced age does not correspond to the quality of the evidence provided. The recommendation 6.30 is based on a Cochrane meta-analysis (one RCT for the endpoint) (Mochtar et al., 2017), the ESPART study (Humaidan et al., 2017) and a study	After extensive discussion, the GDG could not reach consensus with regards to rFSH+rLH and decided not

				by Vuong et al. (Vuong et al., 2015). In Appendix 2 table 22a-b, pages 16 and17, summary of findings, of these guidelines, the results are reported to be based on low/very low-quality evidence because of (1) a serious risk of bias due to poor reporting of methodology, (2) serious inconsistency because only 1 RCT, and (3) small number of events. It is not clear how such poor-quality evidence could lead to a strong recommendation. In addition, grading the recommendation 6.30 as strong is in contrast with the ESHRE Manual for Guidelines Development which states that "When the GDG formulates a strong recommendation, they have to be certain about the various factors that influence the strength of a recommendation" (page 29 of the ESHRE Manual for Guidelines Development). Proposed change of strength of recommendation and quality of evidence: The strength of your recommendation, concerning the use of LH in low responders and advanced age women, is not supported by the quality of evidence. In addition, we suggest splitting the recommendation 6.30 in two different statements.	to formulate any recommendations on the topic.
6	Carlo Alviggi	62-63	1553-1588	The wording of the recommendation concerning low response women may be misleading. Over the last 5 years, 5 meta-analyses have evaluated if supplementation of follicle stimulating hormone (FSH) with luteinising hormone (LH) for controlled ovarian stimulation (COS) might improve ART outcomes (Hill et al., 2012; Lehert et al., 2014; Mochtar et al., 2017; Santi et al., 2017; Alviggi et al., 2018; Conforti et al., 2019). We can conclude from these five meta-analyses that: 1. In the general population (including POR patients), all meta-analyses published in the last 5 years suggested superiority of LH+FSH compared to FSH with respect to ; clinical pregnancy rate (Lehert et al., 2014; Mochtar et al., 2017; Santi et al., 2017), ongoing pregnancy rate (Mochtar et al., 2017), and live birth rate (Lehert et al., 2014). Although there are differences in the statistical significance of the pooled estimates across these papers, there is consistency in the direction of the effect. 2. In the population described as POR (including also suboptimal responders and possibly other subpopulations), all meta-analyses published in the last 5 years have shown that LH+FSH is superior to FSH with respect to clinical pregnancy (Lehert et al., 2014), ongoing pregnancy rate (Mochtar et al., 2017), and live birth rate (Lehert et al., 2014; Mochtar et al., 2017). Although live birth rates (LBR) were not increased after r-hLH supplementation in a large randomized controlled trial in patients with poor ovarian response (POR) defined according to modified ESHRE Bologna criteria, post hoc analysis indicated a possible benefit of r-hLH supplementation to FSH on LBR in a	The existence of multiple meta-analyses on the same research question should not be considered equal to the existence of multiple RCTs on the same research question. The meta-analysis by Lehert et al (2014) has been excluded for reasons mentioned in the guideline (see annex 7) while the meta-analysis by Santi et al (2017) included prospective, longitudinal, and controlled clinical trials and not strictly RCTs. The meta-analysis by Hill (2012) was identified but replaced by the most recent meta-analysis by the Cochrane (Mochtar et al 2017) which was used for this guideline. The study by Alviggi et al (2018) is inaccurately referred to as meta-analysis since it is a systematic review only and was published after

				subgroup of patients with moderate-to-severe POR (Humaidan et al., 2017; Lehert et al., 2018), thus suggesting that the role of LH supplementation in this specific subgroup deserves further investigation. 3. Better defined subpopulations: In suboptimal responders, LH+FSH group was superior to the FSH group concerning implantation rates, clinical pregnancy rates, and LBR (Conforti et al., 2019). In women aged 35-39, LH+FSH group was superior to the FSH group concerning implantation rates, but not for clinical pregnancy rates or LBR (Hill et al., 2012; Alviggi et al., 2018). Thus, we respectfully disagree with the Guidelines Authors that there is sufficient evidence against LH supplementation in low responders. On the contrary, the meta-analyses by Lehert (Lehert et al., 2014) and Mochtar et al. (Mochtar et al., 2017) have shown a clear benefit of LH supplementation concerning critical reproductive endpoints. A) Proposed change of wording of recommendation 6.30: In order to avoid misleading interpretations in this context, the wording of the recommendation should be more precise and explicitly define low responder patients as “poor responders according to ESHRE Bologna criteria”. Justification: A more precise wording is important to avoid confusion with other subgroup populations, such as the hypo-responders, which could benefit from rLH supplementation (Ferraretti et al., 2004; De Placido et al., 2005). The hypo-responder subgroup was first introduced by the Evian Annual Reproduction (EVAR) Workshop Group in 2008. These women had a stagnant response to exogenous FSH during ovarian stimulation and differed from Bologna criteria poor responders in terms of age and ovarian reserve (Alviggi et al., 2018). The role of rLH supplementation in these women has been investigated in two RCTs (Ferraretti et al., 2004; De Placido et al., 2005) which demonstrated that supplementation with rLH significantly improved implantation rate and LBR compared with the rFSH alone regimen. B) Proposed change of strength of recommendation 6.30 (on POR): Given the above considerations and the low quality of the evidence from the two papers used to develop the recommendation, we believe that a statement against the use of rLH in low prognosis women is not supported by strong evidence (see our first comment above).	the guideline search was completed. The study by Conforti et al (2019) was published after the search for the current guideline was completed and examined the addition of LH to rFSH in hyporesponders, not low responders, a population not considered in this guideline due to the inability to define it. Thus, an attempt to derive conclusions, of descriptive nature, from the above studies is probably not going to be useful.

6	Carlo Alviggi	62-63	1553-1588	The inclusion of advanced aged women and low response women in the same recommendation fails to take account of the different grading of the evidence and strength of the recommendation between the two types of patients We strongly believe that -based on the evidence currently available- advanced age patients and the so-called "low response" subpopulations should not be included in the same recommendation. On one hand, there is limited evidence suggesting that rLH supplementation does not offer significant clinical benefit to POR women according to ESHRE Bologna criteria (Humaidan et al. 2017). On the other hand, there is evidence from RCT supporting the use of ovarian stimulation with rLH+rFSH in women aged 35-39 years (Bosch et al. 2011). The recommendation concerning advanced age women is based on one Cochrane meta-analysis and a single RCT (Vuong et al., 2015; Mochtar et al., 2017). The meta-analysis by Mochtar et al. included women ≥36 years old whereas the study by Vuong et al. included women of 35 years and over. We feel there is a discrepancy in the studies included because the study of Matorras et al., which addressed the same topic and included women ≥35 years was not considered (Matorras et al., 2009). Furthermore, we question the inclusion of women beyond 40 years of age. Indeed, the LBR ranges from 31.9% to 22.1% in women aged between 35-39 years undergoing IVF, whereas these rates are below 5% in counterparts aged 40 years and over (Asrm, 2014). It is well know that embryo euploidy rates differ remarkably between women aged 35-39 years and those aged 40 and above (Ata et al., 2012; Asrm, 2014; Esteves et al., 2019). Consequently, women should be stratified in narrower age ranges to properly evaluate the effect of rLH. Notably, rLH supplementation was shown to be beneficial in terms of live births and implantation rate in women aged between 35 and 39 years (Matorras et al., 2009; Bosch et al., 2011; Behre et al., 2015). By contrast, the beneficial effect of rLH supplementation was not shown in studies that included women beyond the age of 40 years (Vuong et al. 2015; Konig et al. 2013, Barrentexea et al. 2008). It is, therefore, possible that the findings by Mochtar et al. (2017) and Vuong et al. (2015) were influenced by the age-related IVF prognosis in their study population. In line with this hypothesis are the results of the meta-analysis by Hill et al., which indicated that women subjected to ovarian stimulation with FSH supplemented with rLH achieved higher clinical pregnancy rates than those with rFSH monotherapy (Hill et al., 2012). The meta-analysis by Hill et al. included mainly women range between 35-39 years old, with the exception of Barrentexea et al. study. Taking together, we strongly believe that the evidence from RCTs does not support a recommendation against stimulation with	After extensive discussion, the GDG could not reach consensus with regards to rFSH+rLH and decided not to formulate any recommendations on the topic.

				rLH+rFSH in women 35-39 years old. Along the same lines, the timing of rLH administration in GnRH antagonist regimens might also explain the discrepancy in findings among the studies. Indeed, in women aged between 35-40 years, rFSH plus rLH from day 1 resulted in a higher benefit in terms of implantation rate, clinical pregnancy per started cycle and clinical pregnancy per embryo transfer than did rFSH on days 1-5 followed by rFSH plus rLH from day 6 onwards (Behre et al., 2015). Other studies confirmed that rLH at the beginning of stimulation (Bosch et al., 2011) was more effective than the addition of rLH during the course of ovarian stimulation in women co-treated with the antagonist regimen (Konig et al., 2013; Vuong et al., 2015). These findings suggest that the combination of rLH+rFSH from day 1, rather than rLH supplementation from day 5 or later, may improve ART outcomes for advanced aged women. Proposed changes of recommendation 6.30 (advanced age): We suggest that a more detailed stratification of advanced age women should be considered in the wording of recommendations to account for the significantly different outcomes between the two groups. The evidence from RCTs does not support a recommendation against ovarian stimulation with rFSH+rLH in women 35-39 years old. Consequently, the recommendation about rFSH+rLH in POR and advanced age should be split in two.	
6	Carlo Alviggi	62-63	1553-1588	Recommendation 6.29: "The addition of recombinant LH (rLH) to recombinant FSH (rFSH) is probably not recommended for controlled ovarian stimulation in the general IVF/ICSI population" The use of live birth as a primary endpoint may not have sufficient statistical power to develop recommendations with a high quality of evidence We agree with the expert committee that cumulative LBR and LBR are the most critical outcomes when evaluating therapeutic strategies in IVF. Nevertheless, other essential endpoints, namely the ongoing pregnancy rate should be considered in studies with low sample size. We feel it is important to report and interpret the findings of the meta-analyses and RCTs in this field with respect to the continuum of reproductive outcomes like implantation rates, pregnancy rates, clinical pregnancy rates, ongoing pregnancy rates, and live birth rates (LBR) (Martins et al., 2018). As a means to avoid publication bias, the effectiveness of an intervention should always consider all available evidence, which is only possible when the most meaningful reported outcome of every published	The choice of critical outcome measures was made a priori, was agreed within the guideline group and represent the outcomes the clinician and the patients are more likely to be interested in. Additional outcome measures were also considered (see introduction) but priority was given, in order to arrive at a recommendation, to the critical outcome measures.

				study is evaluated (Martins et al., 2018). While LBR is the preferred outcome, this endpoint has been reported in only a small proportion of studies. Thus, the estimation of the overall effect of an intervention in systematic reviews using only LBRs might create a biased and partial view of the available evidence (Martins et al., 2018). Indeed, most researches/papers still report intermediate pregnancy outcomes (such as Clinical Pregnancy Rate or Ongoing Pregnancy Rate), which provide relevant information concerning the clinical utility of treatments in reproductive medicine (Braakhekke et al., 2014; Mol et al., 2018). In the Cochrane meta-analysis by Mochtar et al. (2017), the LBR was reported in only four studies that included few patients (4 RCTs, for a total of 499 participants, 53 live births versus 43) treated with rFSH+rLH versus rFSH alone. By contrast, the ongoing pregnancy rate was evaluated in 19 RCTs for a total of 3,129 patients (Mochtar et al., 2017). Notably, the Cochrane meta-analysis by Mochtar considered the quality of the evidence for LBR to be very low, whereas it was judged as "moderate" concerning the ongoing pregnancy rate (Mochtar et al., 2017). Likewise, the difference in sample size is also evident concerning these endpoints when evaluating both the low responder and the women with advanced age (Mochtar et al., 2017). The choice of live birth as the endpoint is even more debatable in low responders and advanced age women given the age-dependent miscarriage rate observed in these subgroups. For example, the miscarriage rate in women over 40 years was estimated to be approximately 30% (Asrm, 2014). Not surprisingly, a dramatic drop-out before delivery was observed during trials, and the sample size required to analyze this endpoint is usually economically unsustainable. Consequently, even very large studies, such as those cited in the draft guidelines (Vuong et al. 2015; Humaidan et al. 2017) were not sufficiently powered to detect differences in terms of live births between treatment arms. Specifically, in the studies mentioned above, the power analysis was calculated based on clinical pregnancy rate and the number of oocytes, respectively. Lastly, the process of converting secondary outcomes such as LBR from original trials to primary outcomes in meta-analyses inevitably leads to a Type I error (Kirkham et al., 2010). Furthermore, it should be taken into account that compared with ongoing pregnancy rate, LBR is more prone to biases not related to IVF. For instance, intrauterine foetal death after 12 weeks of gestation occurs in about 5% of ongoing pregnancies after IVF, whose risk further increases for women of advanced age (Clarke et al., 2010). Given the above, we believe that the use of ongoing pregnancy rate as the main outcome could be more reliable than LBR in this complex context.	

				Proposed change of wording of the recommendation: We believe that the wording would be more precise and useful for the clinician if it stated the endpoints considered, thus "addition of recombinant LH (rLH) to recombinant FSH (rFSH)" does not increase the rate of live birth in the general IVF/ICSI population, but is probably recommended to increase the mean number of oocytes, cumulus–oocyte complexes and embryos and to improve ongoing pregnancy rates	
6	Thomas D'Hooghe *	App.8 p78	Evidence table 6.1.1	The Table of evidence 6.1.1 incorrectly reports the results of the study by Figen Turkcapar et al. In fact, the manuscript authors report that hMG (and not rFSH) was associated with significantly fewer oocytes (9.54 ± 4.31) vs (13.60 ± 5.56, p=0.002), and MII oocytes (7.65 ± 3.39) vs (11.20 ± 5.06, p=0.003) in PCOS patients.	Adjusted.
6	Thomas D'Hooghe *	60	1501	Section 6.1.1 corresponding to Recommendation 26 / 6 Consider rewording the recommendation to include the outcome considered, and to include another recommendation. 6 Proposed wording a) The use of recombinant FSH (rFSH) and human menopausal gonadotropin (hMG) for controlled ovarian stimulation is equally recommended with respect to comparable LBR/ongoing pregnancy. b) However, the use of rFSH is associated with a lower drug usage and significantly higher production of leading follicles, oocytes and MII oocytes than hMG in patients either with or without PCOS. Proposed level of evidence (a) Moderate (b) Moderate Proposed strength of Recommendation (a) Strong (b) Strong Justification - Two of the studies considered (Figen Turkcapar et al., 2013; Ye et al., 2012) demonstrated a significantly higher production of follicles and oocytes in patients allocated to rFSH versus those allocated to hMG, and both trials included cycle characteristics as their primary outcome and LBR as secondary outcome. - The meta-analysis by Lehert et al. (2010), demonstrated that treatment with hMG results in the production of fewer oocytes (-1.54; 95% CI: -2.53 to -0.56; P < 0.0001) compared to rFSH, and that a higher total dose of hMG was administered (MD, 235.46 IU [95% CI: 16.62 to 454.30; P = 0.03]). These findings support the use of rFSH over	The meta-analysis by Lehert was excluded for reasons stated in the guideline (see Annex 7). In addition, because of the availability of RCTs and meta-analyses, retrospective chart reviews were not considered of adequate quality to be considered in formulating guideline recommendations. The choice of critical outcome measures was made a priori, was agreed within the guideline group and represent the outcomes clinicians and the patients are more likely to be interested in. Additional outcome measures were also considered (see introduction) but priority was given, in order to arrive at a recommendation, to the critical outcome measures.

				hMG due to a greater production of oocytes, a similar clinical efficacy and lower costs. - A large retrospective chart review with over 30,000 IVF cycles demonstrated a significantly lower drug usage per cycle for rFSH than for hMG-HP (2072.53 +/- 76.73 IU vs. 2540.14 +/- 883.08 IU, 22.6% higher for hMG-HP; p < 0.01) (Trew et al., 2010). Since the CDG reported the number of oocytes retrieved and the number of MII oocytes retrieved ("yield") among the secondary outcomes used to assess efficacy (see page 9 of the GL document), the CDG should consider including a sentence in the text reporting the superiority of rFSH in follicle and oocyte stimulation versus hMG-HP. The new recommendation should be included based on the res	
6	Thomas D'Hooghe *	63	1580	Section 6.1.4 corresponding to Recommendations 29/ 6 Consider rewording the recommendation to include the outcomes. Proposed wording The addition of recombinant LH (rLH) to recombinant FSH (rFSH) for controlled ovarian stimulation does not increase the rate of live births in the general IVF/ICSI population, but is probably recommended to increase the mean number of oocytes, cumulus–oocyte complexes and embryo quality and to improve ongoing pregnancy rates. Proposed level of evidence Moderate Proposed strength of recommendation Conditional Justification - The Cochrane Review by Mochtar et al. (2017) concluded that the evidence available is insufficient to encourage or discourage stimulation regimens that include rLH combined with rFSH in IVF/ICSI cycles, and reported a VERY LOW quality of evidence with significant heterogeneity among studies I2 = 63% (Mochtar et al., 2017). However, the GDG incorrectly reported a MODERATE quality for the evidence in the Summary of Findings (Table 21, Page 16). See also Mochtar 2017, page 5. The Mochtar (2017) meta-analysis also showed that a combination of rLH + rFSH probably improves ongoing pregnancy rates compared to rFSH alone (OR 1.20, 95% CI 1.01 to 1.42; participants = 3129; studies = 19; I2 = 2%) with MODERATE quality evidence with the addition of about 6% in the ongoing pregnancy rate. - Why is ongoing pregnancy rate not reported in the Summary of Findings table or in recommendations? This outcome was not considered by the GDG even though it is based on more studies (19 RCTs vs 4 RCTs) and a larger study population (3129 vs 499 patients) compared to LBR, with no additional side effects. - The GDG should also consider that, in the general ART population, all meta-analyses	The choice of critical outcome measures was made a priori, was agreed within the guideline group and represent the outcomes clinicians and the patients are more likely to be interested in. Additional outcome measures were also considered (see introduction) but priority was given, in order to arrive at a recommendation, to the critical outcome measures. The existence of multiple meta-analyses on the same research question should not be considered equal to the existence of multiple RCTs on the same research question. The meta-analysis by Lehert et al (2014) has been excluded for reasons mentioned in the guideline (see annex 7) while the meta-analysis by Santi et al (2017) included prospective, longitudinal, and controlled clinical trials and not strictly RCTs. There was indeed a mistake in the strength of the evidence in the Summary of Findings table, which is now corrected.

				published in the last 5 years have shown a significant superiority of LH+FSH over FSH with respect to pregnancy (Santi et al., 2017), clinical pregnancy rate (Lehert et al., 2014; Mochtar et al., 2017), ongoing pregnancy rate (Mochtar et al., 2017) and live births (Lehert et al., 2014).	
6	Thomas D'Hooghe *	Rec 26-32		in our opinion, there is an important imbalance in the wording of recommendations 26 to 32 comparing different gonadotrophin types. According to the outcome data considered by the GDG OS, there are no differences between the intervention and comparison groups for the LB OUTCOMES evaluated based on the data analysed for ALL recommendations 26 to 32. Then, logically, the wording should be consistent across recommendations 26-32. However, this is not the case. On the one hand, the 2 gonadotrophin types evaluated in recommendations 26, 27, 28, and 31 are both equally recommended as the LBR were similar between the 2 groups. On the other hand, other wording is used in recommendations 29, 30 and 32: although the LBR was similar between combination therapy rFSH+rLH and rFSH alone (rec 29 and 30) or to HMG (Rec 32), the recommendation was not that both gonadotrophins are equally recommended, but that the combination therapy rFSH+rLH was not recommended (rec 30) or probably not recommended (rec 29) when compared with rFSH alone (rec 29 and 30) or probably not recommended when compared to HMG (Rec 32).	After extensive discussion, the GDG could not reach consensus with regards to rFSH+rLH and decided not to formulate any recommendations on the topic.
6	Thomas D'Hooghe *	63	1581	Section 6.1.4 corresponding to Recommendations 30/6 Consider dividing the recommendation in two. Proposed wording a) The addition of recombinant LH (rLH) to recombinant FSH (rFSH) is probably recommended for controlled ovarian stimulation in low responders. b) There are not enough data to recommend in favor or against the addition of recombinant LH (rLH) to recombinant FSH (rFSH) in women of advanced age. Proposed level of evidence (a) Low-to-moderate (b) very low Proposed strength of Recommendation (a) Conditional (b) Conditional Justification The GDG made a STRONG a recommendation from a very low-quality evidence. As mentioned the criteria should be justified in the GL text according to GRADE principles.	After extensive discussion, the GDG could not reach consensus with regards to rFSH+rLH and decided not to formulate any recommendations on the topic.

				Indeed, the quality of evidence considered by the GDG is different for the two subpopulations. In particular: - Low responders: The level of evidence derives from 2 RCTs that report opposite results in women with a low ovarian response (Ferraretti et al., 2014; Humaidan et al., 2017). Although the sample sizes are different, the results are inconsistent, and by definition, in this situation any estimate of effect is very uncertain. A STRONG recommendation could be given only in case of a clear advantage in the administration of one of the two options, such as severe side effects or elevated costs or preference of patients. In this case, no difference in cancelled cycles, no cases of OHSS or other side effects are reported in the studies considered. In contrast, a combined protocol (rLH + rFSH) might reduce the rate of cancellation due to imminent OHSS according to data reported in the Cochrane meta-analysis (Mochtar et al., 2017). Even if it is not considered an advantage from a statistical point of view, it cannot be transformed into a disadvantage. - Women of advanced age: The two studies cited reported opposite results (Rahman et al., 2017; Vuong et al., 2015). It is not clear how the GDG formulated a strong recommendation. We therefore propose to split the indications for the management of low-responder women and those with advanced age in two recommendations. Evidence for recommendation (a) LOW/HYPO RESPONDERS In the population described as low/hypo responders, all meta-analyses published in the last 5 years have shown that LH+FSH is superior to FSH with respect to clinical pregnancy (Lehert et al., 2014) or ongoing pregnancy rate (Lehert et al., 2014; Mochtar et al., 2017) or live birth (Lehert et al., 2014; Mochtar et al., 2017). The meta-analysis by Lehert et al. (2014) found a significant benefit for poor responder women receiving rFSH +rLH rather than FSH alone on: - the clinical pregnancy rate (14 studies, n = 1179, RR 1.30 (95% CI 1.01–1.67) - ongoing pregnancy rate (11 studies; 1043 patients) (RR 1.36; 95% CI 1.04–1.79) The subgroup analysis according ovarian response in the Cochrane meta-analysis by Mochtar et al. (2017) reported: - Ongoing pregnancy by ovarian response (19 studies; n 3129, OR 1.20 CI 95% 1.01, 1.42). Studies restricted to women with low response (3 studies; n= 276, OR 2.06, CI 95%1.20, 3.53) These results might be graded low (or moderate) because they come from sub-group analysis of the included trials, but should be considered in the text. Evidence for recommendation (b) ADVANCED AGE	

				- The Mochtar et al. meta-analysis considered women ≥36 years old, therefore studies including women from the age of 35 years were excluded (Matorras et al., 2009), but the GDG included also a study on women ≥ 35 years (Vuong et al., 2015). Indeed, including women beyond 40 years of age is questionable since the rate of live births is below 5% in such patients (ACOG, 2014). Furthermore, also the embryo euploidy rates differ markedly between women between the age of 35-39 years and those > 40 years (ACOG, 2014; Ata et al., 2012). - rLH supplementation in women between the age of 35 and 39 years improved both LBR and implantation rate in three RCTs (Behre et al., 2015; Bosch et al., 2011; Matorras et al., 2009). This was not observed in studies that included women beyond the age of 40 years (Barrenetxea et al., 2008; König et al., 2013; Vuong et al., 2015). Therefore, we believe that the evidence from RCTs does not support a recommendation against stimulation with rLH+rFSH in women 35-39 years old.	
6	Thomas D'Hooghe *	64	1607	Section 6.3 corresponding to recommendation 32/6 – Consider to remove recommendation 32 Justification The study by Pacchiarotti et al. (2010) is of very low quality (as appropriately graded by the CDG), the study sample size was not powered for side effects. The CDG itself expressed some concern about the protocol because "The amount of FSH units required is not compatible with the duration of stimulation and the fixed dose used in both arms". (Annex 8, p. 94) It is difficult to make any recommendation based on such a biased study.	In a small RCT including 122 patients undergoing controlled ovarian stimulation with GnRH agonists, use of rFSH+LH was not associated with increased pregnancy rate compared to hMG (28.3% (15/53) vs. 29.3 (17/58)). However, significantly more cycles were cancelled to prevent OHSS in the rFSH+LH group compared to the hMG group (11.1% (7/53) vs. 1.7% (1/58)) (Pacchiarotti, et al., 2010). The recommendation can be made, but it should, and this is indeed the case, include uncertainty in its formulation.
11	Thomas D'Hooghe *	89	2468	Section 11.3 corresponding to Recommendation 52 / 11 Consider rewording Proposed wording It is not clear whether the addition of a hormonal panel including oestradiol, progesterone and LH measurements is beneficial in terms of efficacy and safety over monitoring by ultrasound alone. Proposed level of evidence Very low	Due to the drawbacks identified by the guideline group in the published studies the current recommendation is a conditional one and incorporates uncertainty, not however ignoring the fact that currently no extra benefit appears to be present by adding E2 assessment to ultrasound

				Proposed strength of Recommendation Conditional Justification Two small RCTs of very low quality support the recommendation formulated by the GDG (two RCTs including only 177 women in total; Golan et al., 1994; Wiser et al., 2012). However, the sample size of both RCTs are too small (Golan et al:114 and Wiser et al: 63) to make any meaningful comparison with statistical power. In fact, there were only 3 and 4 cases of OHSS from both treatment groups in Golan's study and zero OHSS cases in both treatment groups in Wiser's study. Due to a high grade of uncertainty it is hard to make any conclusions on the added value of hormonal monitoring to minimizing OHSS risk. Moreover, there was a similar trend in both studies showing that more oocytes (Golan et al:13.4 vs 11.7; Wiser et al: 11.7 vs 10) and higher pregnancy rate (Golan et al:25% vs 22.2%; Wiser et al: 57.5% vs 40.0%) were achieved in patients who had transvaginal ultrasound (TVUS) plus hormonal assays during ovarian stimulation as compared to those with TVUS only. Even though this trend was not statistically significant different, it is clinically important when developing a recommendation). I Since the available data from the only 2 studies available (Golan et al., 1994; Wiser et al., 2012) are inconclusive. More studies evaluating the optimal procedure for monitoring ovarian stimulation are needed. 10 - Monitoring of oestradiol level is not only essential to minimize the risk of OHSS for patients but also useful for timing the starting dose of GnRH-Antagonist. Orgalutran Summary of Product Characteristics states that "The starting day of Orgalutran is depending on the ovarian response, i.e. the number and size of growing follicles and/or the amount of circulating oestradiol". - Monitoring LH level is also critical for prevention of premature LH surge when such a flexible GnRH-Antagonist protocol is used (Borm and Mannaerts, 2000; Lainas et al., 2005). - Monitoring of progesterone level (at least) on the day of triggering may provide additional information on patient prognosis and facilitate personalized ovulation triggering and embryo transfer strategy. It has been reported that both elevated (Hill et al., 2017; Venetis et al., 2013) and low (Santos-Ribeiro et al., 2014; Vandekerckhove et al., 2014) progesterone levels on the day of hCG are associated with fewer oocytes retrieved and/or reduced pregnancy rate.	for monitoring ovarian stimulation. There is no doubt that monitoring of LH, oestradiol and progesterone levels provides additional information on follicular growth and endometrium status complementing to ultrasonography as the reviewer suggests. However, this was not the question asked in the current guideline. The question was whether this complementary information improves safety and efficacy. Currently this does not appear to be the case. As clearly stated in the guideline it is not clear if this recommendation is valid for patients treated exclusively with GnRH antagonist due to the lack of relevant studies. The association of serum progesterone levels to the achievement of pregnancy has been explored in many studies and meta-analyses and a negative association appears to be present. However, the question examined in the present guideline was whether monitoring of ovarian stimulation by ultrasound and progesterone assessment improves safety and efficacy over ultrasound alone. In this respect no recommendation can be made currently in view of the lack of relevant trials.

				Given the paucity of data and the small population studied, we propose to change this recommendation.	
11	Thomas D'Hooghe *	88	2445	Section 11.1 corresponding to Recommendation 51 / 11 Consider rewording: Proposed wording The addition of oestradiol measurements to ultrasound monitoring is probably recommended. Proposed level of evidence Low Proposed strength of Recommendation Conditional Justification The meta-analysis by Kwan included 6 RCTs and reported broad confidence intervals for most outcomes (Kwan et al., 2014). Kwan et al (2014) concluded that "A combined monitoring protocol including both TVUS and serum estradiol may need to be retained as precautionary good clinical practice and as a confirmatory test in a subset of women to identify those at high risk of OHSS". Accordingly, it is common practice in many countries/regions to have routine monitoring with frequent vaginal ultrasonography and/or serum estradiol measurements in patients with high risk of OHSS, as exemplified by the opinion paper by Humaidan et al. (2016) on reporting of OHSS in clinical trials, and the ASRM guideline on prevention and treatment of moderate and severe OHSS where a 3,500 pg/mL of oestradiol was defined as the cut-off value for OHSS risk. We suggest changing this recommendation due to concerns for patients' safety, in line with the ESHRE guideline development manual.	The question of interest in this case is not whether serum E2 level is associated with OHSS, because this is true. The question of interest is whether the addition of E2 to ultrasound monitoring of ovarian stimulation increases efficacy and safety. In this respect the recommendation made in the current guideline is justified. It is not absolute since the number of patients analysed is not adequate at present and the quality of evidence is low. It is for this reason subject to change in the future if further RCTs show a benefit of E2 assessment in addition to ultrasound evaluation.
13	Thomas D'Hooghe *	94	2609, 2622	Section 13.2 Recommendations 57/13 and Section 13.3 corresponding to Recommendation 58/ 13 Consider removing the recommendations Justification The CDG position on this point is controversial: in the text they state that "The association […] has been studied in several observational studies, but management recommendations cannot be derived from these observational data" (page 94, lines 2616-2618). Despite this fact, the panel formulated a specific STRONG recommendation for clinicians, without, moreover, providing any study in support, a Summary of Findings Table or Table of Evidence. Given the lack of evidence, no	These recommendations were converted to GPP.

				recommendation can be developed. The recommendation as stated, is an opinion of the CDG Overall, the absence of evidence should not result in negative recommendation.	
4C	Thomas D'Hooghe *	12	355	The Lensen Cochrane review cited here was published in 2018 not 2017 (Lensen, S. F., J. Wilkinson, J. A. Leijdekkers, A. La Marca, B. W. J. Mol, J. Marjoribanks, H. Torrance, and F. J. Broekmans. 2018. 'Individualised Gonadotropin Dose Selection Using Markers of Ovarian Reserve for Women Undergoing in Vitro Fertilisation plus Intracytoplasmic Sperm Injection (IVF/ICSI)'. Cochrane Database Syst Rev 2 (February): CD012693. https://doi.org/10.1002/14651858.CD012693.pub2). Throughout the GL text the year of publication seems to be misreported.	Adjusted.
4C	Thomas D'Hooghe *	51	1180	Section 4C.3 corresponding to Recommendation 20/ 4C Proposed wording A gonadotropin dose of 300 IU is probably recommended as a starting dose for predicted low responders, as it may reduce the risk of cycle cancellation. Proposed level of evidence Low Proposed strength of Recommendation Conditional Justification Recommendation 20 is based on a Cochrane meta-analysis with 2 RCTs (Lensen et al., 2018). In the "Summary of findings" of this GL, Annex 2 (Table 13, page 10), the results are based on low quality evidence because of (1) a serious risk of bias due to incomplete reporting of the methodologies in individual studies and (2) the pooled effect included both the line of no effect and the appreciable benefit or harm. The meta-analysis by Lensen et al. (2018), addressed a wide range of dose comparisons, but there were only 2 trials that tested the comparison between 150 IU and 300/450 IU gonadotropins (Klinkert et al., 2005; van Tilborg et al., 2017). An analysis of these two RCT, raises further concerns: 1) The Klinkert and the van Tilborg studies were performed in populations (AFC <5, AFC ≤7, respectively) that did not meet the definition of low ovarian response as defined by the GDG, namely "Low ovarian response is a diminished response to conventional ovarian stimulation, characterized by the presence of a low number of follicles and/or oocytes (Ferraretti et al., 2011). Generally, ≤ 3 follicles on the day of oocyte maturation trigger and/or ≤ 3 oocytes obtained characterize a low response" (see page 9 of these guidelines). 2) Cycle cancellation is an important risk factor that should be considered, from an	The recommendation was changed to 'it is unclear whether a higher gonadotropin dose is recommended over 150 IU. Despite the higher number of oocytes, there was no difference in live birth/ongoing pregnancy rate. Furthermore, the sample sizes of the studies are small and therefore not sufficient to provide evidence for dose comparisons for live birth outcome.

				effectiveness and patient point of view, because one of the major concerns for low ovarian responders is treatment cancellation due to lack of response. 5 It must be noted that 9 patients (34%) in the 150 IU group had dose adjustments due to lack of ovarian response, whereas no dose adjustments were needed in the 300 IU group (Klinkert et al., 2005). In the Van Tilborg et al. 2017 study, in which dose adjustment was not allowed, approximately 30% cycles were cancelled. To conclude, both RCTs on which this recommendation is based, suggest that for predicted low responders, a higher dose of gonadotropins results in more oocytes and reduces the risk of cycle cancellation.	
4C	Thomas D'Hooghe *	51	1181	Section 4C.3 corresponding to Recommendation 21 / 4C Consider to remove the recommendation – or replace it with: Proposed wording “A gonadotropin dose higher than 450 IU is probably not recommended for women with a predicted low response since the balance between potential benefits and harms does not justify its use.” Proposed level of evidence Low Proposed strength of Recommendation Conditional Justification The current recommendation does not seem to be supported by the evidence cited in the GL text. - If the current STRONG recommendation against the use of higher doses (300 IU) is based on the potential risk of OHSS, it should be noted that no events were reported in women receiving 400 IU or 450 IU, and only one event of moderate OHSS was reported in a woman receiving 600 IU (Lensen et al., 2018). The Cochrane meta-analysis by Lensen et al. (2018) reported that “Conclusions about the risk of OHSS are not calculable due to the paucity of events, hence the working group is unable to make any inferences for the outcome of OHSS”. According to the evidence available, the degree of uncertainty is too high to support a STRONG recommendation. Since only one adverse event has been reported in patients allocated to this intervention, and no clear benefits have been reported to support one or the other intervention (only 2 trials), adherence to this recommendation could NOT be used as a quality criterion or performance indicator and it is NOT possible to establish if most	The strength of recommendations is based on a framework, taking into account the evidence, balance between favourable and unfavourable effects and acceptability to stakeholders and patients. Current evidence does not show any benefit of increasing gonadotropin dose beyond 300 IU. In addition, the GDG strongly feels that increasing the gonadotropin dose beyond 300 IU, will only increase patient discomfort.

				patients should receive that intervention or not. Consequently, there is no certainty of the evidence of effects, and a balance between desirable and undesirable effects is not possible. The GDG should recognize that treatment may differ among patients and consider grading the strength of recommendation against the use of higher doses of gonadotropin (450 IU) from STRONG to CONDITIONAL. In addition, the CDG should discuss in the GL text the value patients place on an outcome, the acceptability and feasibility of the intervention, according to the literature or as has emerged in the CDG discussion process.	
	Thomas D'Hooghe *	12	358-359	The sentence "Moreover, a few oocytes more may not make the desired difference in terms of live birth rates" is not supported by a reference. Justification We recognize that the meta-analysis by Lensen et al. (2018) reported a significantly higher production of oocytes in women allocated to higher doses (300/450 IU) of gonadotropin (2 RCTs, MD 0.69, 95% CI 0.5-0.88, 286 women) with no significant difference in live births/ongoing pregnancy rates between the 150 IU and 300/450 IU dose of gonadotropins. Notably, this meta-analysis addresses only the effect of gonadotropin dose changes on number of oocytes and live birth rates and ongoing pregnancy rates, not the direct correlation between number of oocytes and live birth rates/ongoing pregnancy rates. The GDG needs to take into account HIGH, MODERATE and LOW quality evidence supporting a strong correlation between number of oocytes, live births and cumulative live births: • HIGH quality evidence: Toftager et al. (2017). This rigorous RCT with >1000 participants reported a correlation between number of oocytes retrieved and increased chance of live births (aHR: 1.87; 95% CI: 1.13–3.08; P = 0.01 and aHR: 2.15; 95% CI: 1.24–3.71; P < 0.01, respectively). More specifically, a stratification according to number of retrieved oocytes showed an impact on the cumulative live birth rate (LBR), with a "dose"-response gradient varying from 17/82 (20.7%) to 15/58 (43.1%) in the antagonist group (P < 0.01) and from 8/51 (15.7%) to 41/81 (50.6%) in the agonist group (P < 0.01). Those data were confirmed after adjustment for confounding factors. • MODERATE quality evidence: Malchau et al. (2019). Although the CDG would consider studies published up to 8 November 2018 in developing recommendations, we would like to point out that the sentence discussed here is not a recommendation but a statement. According to GRADE methodology, evidence based on observational/cohort studies may be up-graded to MODERATE if based on a very large population (> 30,000 women, Malchau et al. 2019) and/or if they report a dose-response gradient (in	Regarding the correlation between oocyte number and live birth rates, we need to make clear that these cross-sectional correlation data may not automatically imply a causal relationship between oocyte number and live birth rate. The only way to show this is with RCTs, comparing distinct dosage and thereby oocyte number levels. If in this comparison more oocytes in the higher dosage trial arm create more babies, then we have proven that the correlation is indeed without confounding. As of today, none of such studies has ever supported this. Interestingly, the Toftager trial demonstrated that with lower oocyte number, by using the antagonist system, there were not less live births. We have discussed this issue more extensively now in the general introduction.

				Malchau et al. 2019 the odds for live birth was 1.18 [1.07-1.30] for women with 4-9 aspirated oocytes, 1.41 [1.27-1.57] for women with 10-15 aspirated oocytes and 1.63 [1.42-1.88] for women with more than 15 aspirated oocytes). • LOW quality evidence: Drakopoulos et al (2016). This study reports some dose-response gradients, and is on smaller population (i.e., 1099 women). The cumulative LBR significantly increased with the number of oocytes retrieved ($\chi(2)$ test for trend P <0.001). High responders (>15 oocytes) had a significantly higher LBR not only versus poor (0-3 oocytes) (P <0.001) and suboptimal (4-9) responders (P <0.001), but also versus women with a normal (10-15) ovarian response (P = 0.014). Moreover, findings from moderate-low quality evidence retrospective studies on large populations support a positive correlation between the number of oocytes and live birth (Magnusson et al., 2018; Polyzos et al., 2018; Sunkara et al., 2011; Zhu et al., 2018). Overall, this observation is not surprising, as a recently published systematic review has shown that the number of oocytes is positively correlated with the number of good quality embryos (D'Hooghe et al, 2018). Given the evidence available, the sentence should be reworded. Proposed rewording: There is a strong and positive correlation between the number of oocytes retrieved at egg retrieval and (cumulative) live birth rates.	
	Thomas D'Hooghe *	9	259-270	GENERAL REMARKS 1) In the Introduction (page 9), the Guideline Development Group (GDG) defined critical and secondary outcomes for efficacy and safety. Almost all the recommendations include critical outcomes, and secondary outcomes of efficacy are usually not reported in the recommendations in this guideline (GL). This is plausible if the supporting evidence is of high or medium quality. However, when recommendations are based on evidence with a high grade of uncertainty (low or very low quality evidence), the use of secondary outcomes such as clinical pregnancy rate per started cycle, number of oocytes retrieved, and number of MII oocytes retrieved may be useful for clinical decisions. In addition, many RCTs included in the Tables of Evidence consider those outcomes as their primary outcomes. Therefore, the GDG should consider to include both critical and secondary outcomes in the recommendations. 2) Since many different outcomes have been considered in the guidelines, it would be more correct from a methodological point of view, and more helpful from a clinical point of view, to include within the recommendation the specific outcome(s) the recommendation was based on, e.g., "Intervention X is recommended to… [improve the live birth rate] / [increase the number of oocytes]".	Thank you for these comments. 1) Regarding the critical outcomes, it was decided, according to the GRADE methodology, to focus on the critical outcomes first and look at other outcomes when no information is available on the critical outcomes. This makes the process more straightforward and easier to summarize. However, for specific questions, secondary outcomes were considered, and this is reflected in the summary of evidence. 2) We have added this information to the recommendations where relevant. 3) This is correct, and the considerations from the guideline

				3) We are aware that one of the strengths of the GRADE system (compared to previous grading systems) is that high-quality evidence doesn't necessarily imply strong recommendations. Strong recommendations can arise also from low quality evidence. In the present guideline, STRONG recommendations are supported by evidence of low (9 recommendations) or even very low quality (11 recommendations). In all such cases, in which, according to the GRADE system, there is a close or uncertain balance between desirable and undesirable effects, and the "Confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect" (Balshem et al., 2011), the CDG should report in the text the decision making process that led to a STRONG recommendation and should highlight the factors that led to the decision (i.e., patients' perspective, cost-efficacy balance, high consistency of results in many trials, and dose-response effects). Otherwise, we suggest that the GDG reconsider the grading and/or content of the recommendation (Andrews et al., 2013). 4) The CDG decided to use a consistent terminology to report the strength of recommendations throughout the document, i.e., "an intervention is (or not) recommended" for STRONG recommendations and "is (or not) probably recommended" for WEAK/CONDITIONAL recommendations. This choice may be confounding in case of non-inferiority trials in which the evidence indicates no advantage of treatment A over treatment B. Indeed, if no additional side effects, costs or patients' preferences are reported for a given intervention, it does not mean that an intervention should be considered a bad therapeutic option, and it might be incorrect to NOT recommend it. The absence of greater efficacy should not be translated into a recommendation against a specific intervention. 3 5) Similar to what is set out in point 3, the lack of evidence about a specific topic should not be translated into a negative recommendation. In such cases, the GDG may decide not to develop recommendations on that topic.	group leading to the decision of a strong recommendation based on low quality evidence should have been addressed in considerations section. We have checked and added this information where needed. 4) Again, this is a valid point which highlights something the GDG struggled with while applying the GRADE terminology to interventions that are equally regarding efficacy and safety. We have checked and corrected where necessary. 5) For some interventions, the GDG can indeed decide not to make a recommendation. In this guideline, mostly this applies to adjuvant treatments where in the absence of benefit, the GDG decided to recommend not adding additional treatments, rather than refraining from formulating recommendations. This was decided specifically for each intervention and therefore was not further adapted.
9	Filippo Ubaldi	80-81	2145-2172	The growing knowledge of human ovarian follicular waves introduced new models to describe folliculogenesis. This concept has opened a new scenario in which non-conventional COS represents new and intriguing opportunity to fully exploit the waves of human follicular development and to maximize the utilization of the ovarian reserve via tailored protocols especially in very poor prognosis patients. In this scenario, Dual Stimulation (follicular and luteal stimulation) in the same ovarian cycle should be considered a clinical evolution of random start and luteal phase stimulation in order to collect a higher number of oocytes and obtain a adequate number of embryos in all situations where the time is limited and entail non-transfer cycle.	We have to look for the best evidence and as mentioned, there is no randomized trial (neither, considered in second hand, retrospective one) comparing dual stimulation with two conventional stimulations in term of the efficacy (cumulative live birth rates or at least number of oocytes) or efficiency

				We recognize that there are no prospective randomized trials (RCT) that compare dual stimulation with two conventional stimulations in terms of efficacy (cumulative live birth rates) or efficiency (reduced time to live birth) of the two strategies. We also recognize that mandatory freeze-all of oocytes or embryos may be a disadvantage of this protocol because of additional procedure and oocyte manipulation, which, may not be allowed by some national health care. Nevertheless, it must be noted that freeze-all is mandatory also in case of luteal phase stimulation-only, random start, oocyte/embryo accumulation through sequential conventional stimulations and blastocyst stage PGT-A cycles. In addition, we do not understand why Committee Members did not mention in the evidence section that, according to all the papers published on the topic, the mean number of oocytes retrieved in the luteal phase stimulation is significantly higher than follicular phase as are the mean number of blastocysts and of euploid blastocysts. Moreover, the chance to find an euploid embryo or a blastocyst to transfer is significantly higher per started ovarian cycle in the dual stimulation if compared to standard stimulation (Ubaldi et al., 2016, Cimadomo et al., 2018, Vaiarelli et al., 2018). Finally, dual stimulation is applied successfully by many centers in different countries. And the evidence published in favor of this procedure in increasing day by day (Xu and Li, 2013, Kuang et al., 2014, Moffat et al., 2014, Ubaldi et al., 2016, Wei et al., 2016, Tsampras et al., 2017, Vaiarelli et al., 2017, Cardoso et al., 2017, Liu et al., 2017, Rashtian and Zhang, 2018, Zhang et al., 2018, Madani et al., 2018, Jin et al., 2018, Vaiarelli et al., 2018, Alsbjerg et al., 2019, Sighinolfi et al., 2018, Vaiarelli et al., 2019). While we understand that this procedure cannot be suggested for standard patients, poor prognosis patients (e.g., with reduced ovarian reserve, AMA, Bologna POR) or women deserving fertility preservation (oncologic patients) might benefit from it. Hence, confining this technique “only for research” does not reflect the available evidence and could have serious consequences in case of reimbursement or clinical complication In conclusion, although there are no RCTs that show the superiority of dual stimulation vs conventional stimulation in terms of efficacy and efficiency, the author of this guideline could not ignore and/or underestimate the available evidence. We believe that there are enough clinical data to state that “dual stimulation can be considered in poor prognosis patients when freeze-all is mandatory. It is not clear why the Committee stated that: “Luteal phase stimulation could be used in the non transfer cycles” although it has far less clinical and laboratory evidence (some of which use data from dual stimulation) reported in the literature.	(reduced time to live birth). The GDG agrees that in case of urgent fertility preservation (see this section), the dual stimulation, when it is possible, is the can be an option to get more eggs in a shorter time. However, in the dual stimulation there is a proportion of patient that can get an ongoing pregnancy in the first cycle and don’t need the expense of a second cycle.

4C	Fei Gong	51	1180	The statement "A higher gonadotropin dose of 300 IU is probably not recommended over the conventional dose of 150 IU for predicted low responders." is inconsistent with the justification given in the same line which says, "A higher gonadotropin dose of 300 IU daily results in a higher number of oocytes in low responders, and more chances of having an embryo for transfer." Referring to the justification on page 51, this statement is made considering that there are insufficient evidences for dose comparisons for live birth outcome due to small sample sizes. But this statement may mislead clinicians into reducing starting dose to 150 IU, which might keep these patients from obtaining more oocytes or embryos that may increase their probability of getting pregnancy or live birth. In an RCT comparing CPR per embryo transfer cycle between different Gn starting dose in which a total of 95 patients of POR were included, the author found significant higher CPR in 300 IU group over 150 IU group. [1] In 2017, another large RCT which include 511 expected poor ovarian responders found that a higher dose of 450/225 IU per day leads to increased number of oocytes, lower cycle cancelation rate due to insufficient follicle growth. [2] Therefore, the recommendation may be adapted to "A higher gonadotropin dose of 300 IU is probably recommended over the conventional dose of 150 IU for predicted low responders to maximize the cumulative chance of live birth."	The recommendation was changed to 'it is unclear whether a higher gonadotropin dose is recommended over 150 IU. Despite the higher number of oocytes, there was no difference in live birth/ongoing pregnancy rate. Furthermore, the sample sizes of the studies are small and therefore not sufficient to provide evidence for dose comparisons for live birth outcome.
6	Yun Lin	63	1580	The addition of recombinant LH (rLH) to recombinant FSH (rFSH) is not recommended for controlled ovarian stimulation in low responders and women of advanced age. Comment: This conclusion is based on a Cochrane meta-analysis that a large RCT with 939 women in total, the age was 18-41 with poor response according to the Bologna criteria. However, as we all know, the reason for poor responder as Bologna criteria for people <35 years old is complex and the definitions of POR were heterogeneous and limited the comparability across the study. Recently, a new systematic review in 2018 which investigated six population group, unlike other studies, the authors draw a distinction between hypo-responders and poor responders and analyzed these groups separately1. The authors concluded that "Recombinant hLH supplementation appears to be beneficial in two subgroups of	After extensive discussion, the GDG could not reach consensus with regards to rFSH+rLH and decided not to formulate any recommendations on the topic. Regarding the systematic review, the reviewer probably refers to the manuscript by Alviggi et al (2018) which however, was not accompanied by a meta-analysis producing effect estimates and thus

				patients: 1) women with adequate prestimulation ovarian reserve parameters and an unexpected hyporesponse to r-hFSH monotherapy; and 2) women 36-39 years of age." Therefore, we should consider the effect of rLH in hypo-responders which are different from the poor responders defined by Bologna criteria. It is possible that such patient population required higher doses of gonadotropins in the second cycle but according "two cell–two gonadotropin'' model LH stimulates theca cells thereby advancing androgen production, and FSH governs the proliferation of granulosa cells (GCs) and promotes E2 synthesis. The rLH supplementation was therefore supposed to be beneficial in hypo-responders. In my opinion, the recommendation should be changed to: The addition of recombinant LH (rLH) to recombinant FSH (rFSH) is probably recommended for controlled ovarian stimulation in hypo-responders and women of advanced age.	is not able to change the recommendations made.
16	Jan Olofsson *	108		Dydrogesterone is probably… Comment: Why probably after 2 large phase 3 studies vs MVP with ~ 2000 patients? Dydrogesterone is now approved in 57 countries for the luteal phase support indication. The correct statement is that dydrogesterone is recommended for luteal phase support. Its efficacy and safety (OHSS) are equal to progesterone. This should also be brought in line with page 23.	The GDG has re-discussed this recommendation, however, considers the safety data from 2000 patients insufficient to make a firm statement and there is a lack of long-term offspring health studies.
16	Jan Olofsson *	109		The evidence suggests that when compared to progesterone, oral dydrogesterone Comment: Why "suggests" after 2 phase 3 studies vs MVP with ~ 2000 patients and a meta-analysis? (References: Tournaye, et al., Hum Reprod 2017;32(5)1019–1027, Griesinger, et al., 201833(12):2212-2221. The correct statement is that dydrogesterone when compared to progesterone has similar efficacy.	Rephrased to: When compared to progesterone, oral dydrogesterone has similar ongoing pregnancy rate.
16	Jan Olofsson *	109		Please also add the following phrase: oral dydrogesterone has similar ongoing pregnancy rate in 2000 patients in two RCTs (Tournaye 2017; Griesinger 2018)	The study by Tournaye is included in the Barbosa (2018) meta-analysis, meaning that the ongoing pregnancy rate is already included in the evidence section. Furthermore, ongoing pregnancy rate is was not defined as a critical outcome by the GDG, meaning it did not influence the recommendation.
16	Jan Olofsson *	109		Please change the sentence to the following: …The studies by Tournaye et al. and Griesinger et al. reported similar safety and tolerability in both treatment groups (Tournaye, et al., 2017, Griesinger, et al., 2018).	Rephrased to: The studies by Tournaye et al. and Griesinger et al. reported similar safety and

					tolerability in both treatment groups (Tournaye, et al., 2017, Griesinger, et al., 2018).
16	Jan Olofsson *	109		As dydrogesterone is a synthetic form of progesterone. Suggest change to: As dydrogesterone is a orally active progestogen, a retroproesterone, with a chemical structure and pharmacological profile that closely resembles that of endogenous progesterone	Rephrased to: As dydrogesterone is an orally-active progestogen different in structure from natural progesterone, …
16	Jan Olofsson *	109		Long-term offspring health studies are currently lacking. Comment: there are no long-term offspring heatlh studies on bioidentical progesterone given at rather high vaginal doses with high exposure to the uterus and early fetus either. However, dydrogesterone was first registered in 1960 in Europe and has a long track record of pregnancy use. It is estimated that 113 million women and about 20 million fetuses have been exposed to dydrogesterone since 1960. There is no pharmacovigilance data indicating an increased risk of the offspring, short-term or long-term.	The GDG has considered your comment, however, has decided to keep the statement as it is. For natural progesterone as well, it is stated in the guideline that long-term offspring health studies are currently lacking.
11	Aidong Gong	94	2622	Recommendation 51: The addition of oestradiol measurements to ultrasound monitoring is probably not recommended. Recommendation 58: It is not recommended to base timing of final oocyte maturation on oestradiol/follicle ratio. There is no data from prospective RCT showing how to associate oocyte maturation with oestradiol/follicle ratio but a retrospective study of 342 in vitro fertilization cycles with normal ovarian reserve in women who underwent long GnRH agonist protocol showed that pregnancy rate is better when E2/fol is between 200 and 299.99 pg/ml. Also, increasing serum E2/fol positively correlates with better oocytes and embryo quality. (J Obstet Gynaecol India. 2014 Apr;64(2):124-9.). It is more reasonable to say “there is a lack of evidence to confirm whether it is beneficial	A recommendation for an intervention is preferably not based on retrospective studies since these are particularly prone to bias. Recommendation 58 was however changed from 'strong' to GPP.
12	Aidong Gong	91	2531	Recommendation 53: Routine monitoring of endometrial thickness during controlled ovarian stimulation is probably not recommended. It is common practice to monitor endometrial thickness for planning whether to go for fresh transfer or freeze-all, and the exact day of embryo transfer. On one hand, a paper from China showed monitoring the endometrial thickness and classification not only can assess the receptive of endometrium, also helps to predict the pregnancy outcome after IVF - ET. (Chin J Fam Plann, V01. 26, No. 10. October 2018). Another Chinese paper showed endometrial thickness during fresh IVF cycles was a better predictor of endometrial receptivity in subsequent FET cycles than FET cycle endometrial thickness. For those females with thin endometrium in fresh	The first publication mentioned is non-English, and the second publication was excluded from the body of evidence because it also includes FET and HRT cycles.

				cycles, additional estradiol stimulation might be helpful for adequate endometrial development (Medicine (2018) 97:4). Thus, monitoring of EMT during COS still provides useful information for patient prognosis. I suggest deleting this recommendation until further data from large RCT showing no value of EMT monitoring during COS.	
6	Pablo Diaz-Spindola	18		In recommendation 30 about adding LH to gonadotropins, and it was "NOT RECOMMENDED" for poor responders and advanced age, We are totally agree about not to use in general population, but in low responders and over 35 yo. Alviggi et al 2018 in a systematic review considered benefits of LH in hiporesponders and women 36-39 yo	After extensive discussion, the GDG could not reach consensus with regards to rFSH+rLH and decided not to formulate any recommendations on the topic. The systematic review by Alviggi et al (2018) was not accompanied by a meta-analysis and thus it did not produce effect estimates that could be used in formulating a recommendation.
11	Pablo Diaz-Spindola	88		In recommendation 51, we consider Oestradiol and ultrasound monitoring is mandatory to avoid OHSS and decide total freeze technique. ASRM 2016 guidelines consider measured and us monitoring to avoid OHSS with the cut-of level of 3500 pg/ml. and Humaidan et al 2016 suggest routine us monitoring and serum estradiol measurements in for IVF treatments. (Oestradiol measurement and ultrasound monitoring is recommended in most of the cases of COH)	A recommendation cannot be based on opinion papers (Humaidan et al 2016) or other guidelines (ASRM). The question of interest in this case is not whether serum E2 level is associated with OHSS, because this is true. The question of interest is whether the addition of E2 to ultrasound monitoring of ovarian stimulation increases efficacy and safety. In this respect the recommendation made in the current guideline is justified. It is not absolute since the number of patients analysed is not adequate at present and the quality of evidence is low. It is for this reason subject to change in the future if further RCTs show a benefit of E2 assessment in addition to ultrasound evaluation.

18	Pablo Diaz-Spindola	121		The freeze all protocol, we need to include the new meta-analysis from Matheus et cols 2018, to know, the real indication for freeze all technique. Of course the most important indication is to avoid OHSS. Besides this, we can increase the indications and recommendations of this technique.	Based on the comment, it is not clear to the GDG which reference is referred to. The GDG has re-evaluated the recommendation and still stands by it.
4C	Pablo Diaz-Spindola	17	R 20	In recommendation 20 we are not agree about it. Papers that support it do not meet the criteria of poor responders. And they do not considered the cancelation rate of patient with 150 UI. La Marca et Al 2018 suggest in poor responders a higher dose of gonadotropins for more oocytes and avoid cycle cancellation. (Probably recommended)	The recommendation was changed to 'it is unclear whether a higher gonadotropin dose is recommended over 150 IU. Despite the higher number of oocytes, there was no difference in live birth/ongoing pregnancy rate. Furthermore, the sample sizes of the studies are small and therefore not sufficient to provide evidence for dose comparisons for live birth outcome.
	Pablo Diaz-Spindola	8	236	In Mexico, as in many latin American countries, we use international GDG´s as ASRM for the COS protocols, and it is very important to clarified terminology. For example, the term "recommended" means for us it is beneficial for the patient, but when we used "NOT RECOMMENDED" it means that it could be harmful for patient. An in legal situation, it could be detrimental for physicians. And most of the times the term is used to define the treatment or medication in not useful for the patient.	We use universally accepted terminology to formulate recommendations in the ESHRE guidelines, according to GRADE methodology. The GRADE methodology is explained in detail in annex 5: methodology.
	Pablo Diaz-Spindola	9	253	I understand that you used Bolonia criteria for define low responders patients. I think you need to extend this definition. Maybe you can used or add POSEIDON criteria for poor responders patients (Alviggi 2016). I know that this is one of the most difficult definitions to clarify in this moment. But maybe we should consider young poor responders patients. WE consider it is very useful.	The GDG is very aware that there are more nuanced definitions of a low ovarian response but decided to try and keep it simple: in the guideline there is only low-normal-high responder. Therefore, the GDG only suggested definitions for low and high response.
	Pablo Diaz-Spindola	9	259	Talking about critical outcome, the GDG has to define it. We know at this moment cumulative live birth rate per started cycle is the most efficacy result. In advance we want to promote in all world SET, freeze all technique for the indicate patient and avoid multiple pregnancy and moderate or severe OHSS.	Cumulative live birth was defined as cumulative live birth per started cycle, fresh+frozen. Embryo transfer will be addressed in a separate guideline.

	Pablo Diaz-Spindola			In summary we consider that it is no easy to make a GDG for Ovulation Induction, but I concern about all research advance up to this moment, that we know we need highest evidence base knowledge. But what about FSH, LH Polymorphism, and number of oocyte we need to get an euploid embryo? Please, consider terminology for most of the languages around the world, because ESHRE is one of the most important and distinguish fertility organization in the world, and we used you GDG´s for patients' treatments.	Ovarian stimulation is a very broad topic for guideline development. The GDG tried to focus on the most important issues, to avoid making the guideline too extensive. Different types of FSH were compared in chapter 6. LH polymorphisms and the 'ideal' number of oocytes were considered outside the scope of the guideline. However, the GDG can consider adding these topics for the update of the guideline in 2 years. ESHRE produced its guidelines in the English language and has a translation policy for national societies who wish to adapt the guideline in their language.
2	Ahmet Turp	34, 35	699, 733	Serum progesterone is sometimes necessary especially woman with age older than 39. These women suffer from poor response and asynchrony of follicle recruitment. If initial progesterone level which is at day 2 is higher than progesterone levels ≥ 90 ng/dL are related to continuing activity of the corpus luteum from the menstrual cycle (lutein unruptured follicle syndrome different diagnosis with large follicles :asyncrony) which may inhibit or delay follicle development and to decide whether to start or not to a good cohort follicle for gonadotropin treatment (1). 1. Glenn L. Schattman, Sandro C. Esteves , Ashok Agarwal Unexplained Infertility Springer; 2015 : Part IV Evaluation p266.	The current evidence on the effect of elevated progesterone levels on cycle day 2 is not solid and the clinical value of the test was not assessed. Until more evidence is available, the GDG thinks that assessment of progesterone level on day 2 of the cycle at the start of ovarian stimulation is probably not recommended
/	Ahmet Turp			The authors can mention some about endometriomas with ovulation induction. (COS+endometriomas). 1.Seyhan A, Urman B, Turkgeldi E, Ata B. Do endometriomas grow during ovarian stimulation for assisted reproduction? A three-dimensional volume analysis before and after ovarian stimulation. Reprod Biomed Online. 2018 Feb;36(2):239-244. 2. Ferrero S, Scala C, Tafi E, Racca A, Venturini PL, Leone Roberti Maggiore U. Impact of large ovarian endometriomas on the response to superovulation for in vitro fertilization: A retrospective study. Eur J Obstet Gynecol Reprod Biol. 2017 Jun;213:17-21.	Endometriomas with ovarian stimulation is considered outside the scope of this guideline.

/	Ahmet Turp			The authors can mention some cyst aspiration prior to ovarian stimulation whether it is preferable or not. (may be under non-conventional start part, or ajjuvant therapies). 1. Levi R, Ozçakir HT, Adakan S, Göker EN, Tavmergen E. Effect of ovarian cysts detected on the beginning day of ovulation induction to the success rates in ART cycles. J Obstet Gynaecol Res. 2003 Aug;29(4):257-61. 2. 2. Management of prestimulation ovarian cysts during assisted reproductive treatments: impact of aspiration on the outcome. 3. McDonnell R, Marjoribanks J, Hart RJ. Ovarian cyst aspiration prior to in vitro fertilization treatment for subfertility. Cochrane Database Syst Rev. 2014;(12)	Treatment of ovarian cysts before ovarian stimulation is considered outside the scope of this guideline.
	Ahmet Turp	12	330	The authors comment on recruiting primordial oocytes to antral follicle takes two months however it takes more than 2 months to get from primordial follicles to pre-antral or antral follicle and it is about 360 days to get visualized by Transvaginal ultrasonography. 'The duration of development from the primary follicles to the secondary follicles is required for about 120 days, and the development from the secondary follicles to the antral follicles is needed for approximately 71 days, whereas only 14 days are inquired to development from the antral follicles to a pre-ovulatory follicle.' 1.McGee EA, Hsueh AJW. Initial cyclic recruitment of ovarian follicles. Endocrine Review. 2000;21 (2):200–14.	The reviewer is correct. This was corrected in the introduction.
	Ahmet Turp	9	268	The authors comment on patients related outcomes which listed some patient related factors. I will recommend adding some terminologies like 'cost of gonadotropins and the time of the injection period (1,2). There is also a terminology that patient friendly ovulation induction protocols which simplifies daily injections. These two factors are very important in ovulation induction protocols. There are many studies that includes cost-effective studies because the cost of ART is very important issue in therapy. Some patients prefer mild stimulation protocols because of the high cost of therapies. 1. Busnelli A, Somigliana E. Prognosis and cost-effectiveness of IVF in poor responders according to the Bologna Criteria. Minerva Ginecol. 2018 Feb;70(1):89-98. 2. Olivennes F. Patient-friendly ovarian stimulation. Reprod Biomed Online. 2003 Jul-Aug;7(1):30-4. Review.	As stated in the scope of the guideline, patient-related outcomes such as compliance and patient preferences were included in the guideline. Costs were not included in the guideline since these may vary substantially between European countries.

1	Tamar Barbakadze	15	1	AFC is the most important and accurace criteria for prediction of ovarian response	Current evidence shows that both AFC and AMH have a high accuracy in the prediction of ovarian response
6	Tamar Barbakadze	18	26	In HMG stimulated cycles there is higher embryo euploidy rate	Embryo euploidy rate is not one of the critical outcomes in this guideline.
8	Tamar Barbakadze	21	38	Testosterone does not change prognoses of ovarian response in law responders	The GDG formulated the recommendation that use of testosterone before controlled ovarian stimulation is probably not recommended for low responders.
8	Tamar Barbakadze	21	39	DHEA does not change prognoses of ovarian response in law responders	The GDG formulated the recommendation that use of DHEA before and/or during controlled ovarian stimulation is probably not recommended for low responders
13	Tamar Barbakadze	21	56	Maturity status of oocytes depends on follicle size	As stated in the recommendation, the decision to trigger final oocyte maturation should take into account several factors, not only follicle size.
18	Tamar Barbakadze	24	86	Freeze all is the safest approach to prevent OHSS	The GDG recommends a freeze-all strategy to fully eliminate the risk of late-onset OHSS.
6	Sandro C Esteves	18	REC 6.26	Recommendation 6.26: "The use of recombinant FSH (rFSH) and human menopausal gonadotropin (hMG) for controlled ovarian stimulation is equally recommended" Strength – Strong Justification: "The results from the meta-analysis suggest a slightly higher efficacy (LBR/PR) with hMG compared to FSH in an GnRH agonist cycle which was not considered clinically relevant, and with no difference in safety, the GDG concluded that hMG is probably not superior to rFSH. This conclusion is supported by the results of studies published after the meta-analysis". Comments: Although this recommendation was ranked as "STRONG", we believe it is not consistent with the ESHRE Manual for Guideline Development, which states that "When the GDG formulates a strong recommendation, they have to be certain about the various	The choice of critical outcome measures was made a priori, was agreed within the guideline group and represent the outcomes clinicians and the patients are more likely to be interested in. Cumulative live birth rate was chosen as a critical outcome for this guideline. Unfortunately, it is still infrequently reported. Heterogeneity, on the other hand, is something inherent when evaluating clinical trials and for this reason several techniques have

				factors that influence the strength of a recommendation". In fact, the evaluated studies supporting this recommendation are very heterogeneous, concerning the inclusion criteria and the study protocol design types. The Cochrane meta-analysis from van Wely et al., evaluated 12 trials, and the down regulation protocols varied among them. Only one of the studies was performed under the GnRH antagonist protocol (Bosch et al., 2008) and to our knowledge this was the first RCT to compare recombinant FSH to hMG in antagonist protocol cycle. Although ongoing pregnancy per started cycle, implantation, clinical pregnancy, and pregnancy loss rates were similar when comparing both gonadotropins, more oocytes were obtained from patients receiving recombinant FSH then hMG (14.4±8.1 versus 11.3±6.0, respectively; P=0.001) (Bosch et al., 2008). Although the authors did not evaluate CLBR, we can hypothesize that this difference in the number of retrieved oocytes might be associated with increased CLBR in recombinant group when compared to hMG. The results presented by Bosch et al., are in contrast to those presented by Devroey et al., which were considered for the present recommendation; this latter study found no differences in clinical outcomes per cycle and also in CLBR (Devroey et al., 2012). It is important to acknowledge that when evaluating the RCTs Considered by the GDG, recombinant FSH resulted in a higher number of oocytes per oocyte retrievel (Bosch et al., 2008; Devroey et al., 2012; Figen Turkcapar et al., 2013) and the use of recombinant gonadotropins was associated with a lower gonadotropins consumption (Devroey et al., 2012) and a shorter treatment duration (Figen Turkcapar et al., 2013) when compared to hMG. Thus, the strength of this recommendation is conditional, and we suggest changing the wording as below. Proposed change of the wording of the recommendation: • The use of recombinant FSH (rFSH) and human menopausal gonadotropin (hMG) for controlled ovarian stimulation is equally recommended concerning LBR per fresh embryo transfer in patients treated with a long GnRH agonist protocol. However, more oocytes can be obtained in cycles with recombinant FSH".	been used for its management when synthesizing data. It should be clear though that although it is not possible to produce a recommendation in the setting of a guideline group based on theoretical arguments, it is possible to arrive at a recommendation in the presence of RCTs. Their quality and their size will define the strength and the uncertainty of a recommendation, as in the current case.
4C	Sandro C Esteves	17	REC 4C.20	Recommendation 4C.20: "A higher gonadotropin dose of 300 IU is probably not recommended over the conventional dose of 150 IU for predicted low responders" Strength – Conditional As stated, this recommendation might be misinterpreted. It seems that there is no benefit in using a dose of 300 IU over 150 IU for predicted low responders. The justification for this recommendation is: "A higher gonadotropin dose of 300 IU daily results in a higher number of oocytes in low responders, and more chances of having	The recommendation was changed to 'it is unclear whether a higher gonadotropin dose is recommended over 150 IU. Despite the higher number of oocytes, there was no difference in live birth/ongoing pregnancy rate. Furthermore, the

				an embryo for transfer. There is unlikely to be significant benefit with dose >300 IU daily". However, this statement is solely based on the OPTIMIST study which refers to CLBR over a period of 18 months rather the CLBR as defined by this guideline. Proposed change of the wording of the recommendation: The wording would be more precise as: Both gonadotropin doses of 300IU and 150IU are equally recommended for predicted low responders, with an advantage of the former in terms of number of oocytes retrieved and cycle cancellation.	sample sizes of the studies are small and therefore not sufficient to provide evidence for dose comparisons for live birth outcome.
	Sandro C Esteves	12		Subsection – Oocyte number and Dosage: What is the relation like? In the last sentence of the first paragraph, it is stated that "a few oocytes more may not make the desired difference in terms of live birth rates." However, this statement is not following the latest evidence, based on large databases. Sunkara et al. evaluated more than 400,000 cycles of IVF and demonstrated that each additional oocyte obtained could increase live birth rates in a fresh embryo transfer cycle, independently on the ovarian response. This study specifically did not consider cumulative pregnancy rates (Sunkara et al., 2011). It has been recently shown by the largest multicenter study evaluating the impact of ovarian response on CLBR, that there is a progressive increase of CLBR with the number of retrieved oocytes and that the ovarian stimulation has no detrimental effect on oocyte/embryo quality (Polyzos et al., 2018). The above-mentioned findings were also corroborated by Malchau et al., that recently evaluated the Danish National IVF-registry and found that the number of retrieved oocytes in the first cycle was associated with CLBR (Malchau et al., 2019). In a recently published Delphi-consensus by Bosch et al., a decision-making analysis were established with the participation of 27 experts from all around the world, some of these that are also the authors of the present ESHRE guidelines. The authors stated that CLBR per cycle, including live births from fresh and frozenthawed embryos, significantly increases with the number of retrieved oocytes (Bosch et al., 2019). McLernon et al., developed a prediction model to estimate the chances of a live birth over multiple complete cycles of IVF treatments, and found that the number of collected eggs was considered a post-treatment predictor of live birth (McLernon et al., 2016). This model was validated by Leijdekkers et al., who performed an external validation study and concluded that McLernon models could accurately predict CLBR (Leijdekkers et al., 2018). The aforementioned data show the importance of the number of oocytes retrieved per cycle. Interestingly, this importance was recognized on Page 17 - 4B – 17 of the present consensus, where the recommendation is that "A reduced gonadotrophin dose is probably not recommended over conventional gonadotrophin dose for predicted	Regarding the correlation between oocyte number and live birth rates, we need to make clear that these cross-sectional correlation data may not automatically imply a causal relationship between oocyte number and live birth rate. The only way to show this is with RCTs, comparing distinct dosage and thereby oocyte number levels. If in this comparison more oocytes in the higher dosage trial arm create more babies, then we have proven that the correlation is indeed without confounding. As of today, none of such studies has ever supported this. Interestingly, the Toftager trial demonstrated that with lower oocyte number, by using the antagonist system, there were not more live births... We have discussed this issue more extensively now in the general introduction.

				normal responders." The justification for this recommendation was "Although available studies suggest similar efficacy in terms of clinical pregnancy rate between reduced-dose and conventional dose stimulation, the lower number of oocytes retrieved could potentially compromise cumulative live birth rate in predicted normal responders", thus emphasizing the importance of the number of oocytes during one IVF cycle. Proposed change of the wording of the sentence: • A few oocytes more might make a difference in terms of live birth rates and cumulative live birth rates.	
	Sandro C Esteves	12		Subsection – Oocyte number and Dosage: What is the relation like? In the second paragraph, in the 3rd line, it is stated "Reduction of the FSH stimulation dosage may bring a more mitigate response, with better safety, without jeopardizing overall live birth prospects. Most of the studies comparing lower doses of gonadotropins with standard dosage were performed in GnRH agonist long protocol with the hCG for triggering the final oocyte maturation. Nowadays, with the use of GnRH antagonist protocols with GnRH agonist for triggering final oocyte maturation, the risk of OHSS is negligible, and such protocols also allows clinicians to recommend elective embryo freezing (freeze-all cycle) in patients with risk of OHSS development. Thus, it is feasible to maximize the number of retrieved oocytes to maximize the CLBR while securing the safety of the treatment (Devroey et al., 2011). Along the same lines, it is possible to perform a safe treatment avoiding OHSS development, without the reduction of the gonadotropin dosage that may have a detrimental effect over the CLBR per started cycle, as the lower the dose of gonadotropin the higher the risk of cycle cancelation due to poor response when comparing to standard dosage (Heijnen et al., 2007; Oudshoorn et al., 2017; Van Tilborg et al., 2017). Proposed change of the wording of the sentence: • Reduction of the FSH stimulation dosage may bring a more mitigate response, with better safety in GnRH agonist protocols. However, it may jeopardize the overall live birth prospects and CLBR per started cycle.	This section has been adjusted.
	Sandro C Esteves			"The critical outcomes for this guideline are efficacy in terms of cumulative live birth rate (CLBR) per started cycle and live birth rate (LBR) per started cycle; and safety in terms of moderate and/or severe OHSS" It is critical that in the recommendations a clear distinction is made between these two outcomes. The fact that some strategy does not prove to be superior to another strategy in one fresh cycle (LBR) does not mean that this strategy is not superior to another one when comparing cumulative rates (CLBR). Thus, for	For very few recommendations CLBR was reported. The recommendations are always based on the critical outcomes, as formulated in the scope section of the guideline. Available evidence regarding the critical outcomes is described in the

				recommendations that did not use publications reporting CLBR, the recommendations need to specify that they apply to fresh cycles only. It is also essential to adequately evaluate the definition of CLBR in different studies. The definition of this Guideline is CLBR per cycle initiated. Thus, studies comparing cumulative rates by period and not by cycle should be considered with caveats. Some studies compare different strategies evaluating cumulative rates by a period (for example after 12 months or 18 months), but not considering the number of cycles necessary to reach this final result. This is probably an issue when considering the OPTIMIST study, in which they compared individualization versus no individualization (fixed daily 150 IU of gonadotrophin) and concluded that there were no differences in CLBR. However, the CLBR was considered per period (18 months) independently on the number of cycles that was necessary to achieve the outcome (van Tilborg et al., 2017).	evidence section of each recommendation.
15	Gemma Castillón Cortés	104	2884	I miss more information about dual-trigger vs double- trigger in low responders or patients with high ratio of inmaturity in COS.	Patients with a high ratio of immaturity in ovarian stimulation is a very specific patient group, which is not in the scope of this guideline. However, we added a statement to the justification.
4	William Ledger			Many papers support individualization of management in these complex cases. Is the guideline advocating slavish adherence to 150 IU for cycle after cycle? This is counterintuitive There are no RCTs that tell us what to do in cycle #4, for example. This must be individualised	Conventional gonadotropin dosing is 150-225 IU. Current evidence in low responders shows that increasing the dose of gonadotropin beyond 300 IU has no benefit on clinical outcomes (Lensen 2017).
6	William Ledger	60-62	/	Evidence for urinary vs rec FSH largely supports equality of efficacy. There have been many industry supported meta analyses which differ in their conclusions. The ovary cant read the the label on the meds - they are all equally efficacious. Industry supported analyses usually support the product of the sponsor and should not be over relied upon	The GDG agrees with your comment, rFSH and all types of urinary FSH were equally recommended.
11	William Ledger	88	2445	Advocating use of US alone for cycle monitoring is a recipe for disaster. Ignoring pre-trigger E2 will lead to more cases of OHSS. Options include cycle cancellation, reduced dose of hCG trigger or agonist trigger in antagonist cycles. OHSS continues to be a lethal condition and such a case would be open to legal challenge as medical negligence if E2 had not been measured ore-trigger eg viz ASRM guidelines and Humaiden papers on avoidance of OHSS.	A high serum E2 concentration on the day of hCG trigger has been suggested as a predictor of OHSS. However, high or rapidly rising E2 levels alone are unreliable and poor predictors of OHSS [Mathur et al

					2000, Alper et al 2009). On the other hand, the number of follicles on the day of hCG was shown to be superior to E2 levels in predicting severe OHSS (Papanikolaou et al 2006, Griesinger et al 2016). The above studies, however cannot justify or deny the use of a specific monitoring protocol for ovarian stimulation. For that purpose, the appropriate studies should be performed.
11	William Ledger	89	2456	Progesterone should also be measured during stim – see many papers on elevated P4 and poor cycle outcome. The decision to freeze all may be best in such cases. Eg viz Kolibianakis, Venetis	The association of serum progesterone levels to the achievement of pregnancy has been explored in many studies and meta-analyses and a negative association appears to be present. However, the question examined in the present guideline was whether monitoring of ovarian stimulation by ultrasound and progesterone assessment improves safety and efficacy over ultrasound alone. In this respect no recommendation can be made currently in view of the lack of relevant trials.
4A	William Ledger	41	1089	The Optimist study was, as its name implies, optimistic. The inclusion and exclusion criteria were very tight but the authors of this guideline have accepted its conclusions uncritically – you cannot extrapolate from "optimist" to all patient groups. You cannot extrapolate its findings to all patient groups. Definition of poor responder was non-standard and it ignored cycle cancellations	The reviewer is probably referring to chapter 4A.2.3, reduced dose protocol in high responders. In the justification it is explained that there are several shortcomings to this study, and caution is necessary when interpreting the results. The strength of the recommendation was however adjusted from strong to conditional.

	William Ledger			The authors of the guideline seem to confuse data from IUI with IVF – targets for stimulation are clearly different and must not be taken together	The GDG confirms that in this guideline, only data from IVF/ICSI studies were used. Furthermore, no targets for stimulation were formulated in this guideline.
	William Ledger			This is a long guideline – the duration of its being open for comments is much too short to allow the IVF community time to digest and respond to it.	We will take your comment in consideration for the next stakeholder review.
	William Ledger			The evidence is frequently over-interpreted, for example in extrapolating from long downregulation protocols to all cycle types and from young normal responder patients to all patient types	The evidence was only extrapolated when there was little evidence available and it was used as indirect evidence.
	William Ledger			Target number of oocytes is also controversial. Some groups (oocyte donors, PGT-D, oocyte freeze patients) should aim for more. Many papers suggest that more may be better when CLBR is used as a denominator. To advocate strongly in favour of a modest target is to over interpret the data.	There is no clear definition of a too high or too low ovarian response. The GDG hopes that this guideline may lead to more studies regarding oocyte numbers.
	William Ledger			This guideline will be used by national funding bodies to restrict our ability to undertake best practice in ART. It is over proscriptive, uses low quality evidence to support didactic recommendations and fails to acknowledge uncertainty and lack of good quality evidence.	The few strong recommendations are clearly based on evidence for efficacy or safety aspects and should be applied to all patients in all conditions. However, the majority of the recommendations are conditional recommendations, which leaves room for clinical interpretation and adaptation to national context. The guideline contains a table on page 114 explaining the difference between strong and conditional recommendations and how this should be interpreted by clinicians, patients and policy makers.
6	Yuan Li	65	1638	Recommendation: "There is no evidence available to recommend the substitution of FSH by Clomiphene Citrate in controlled ovarian stimulation." Comment: There were published evidence showed that concomitant using CC and Gn reduces the	The publication by Satwik and Kochhar 2018 was excluded from the body of evidence because it does not investigate substitution of FSH by

				Gn dose per retrieved oocyte and per good embryo (J Obstet Gynaecol Res. 2018 Jun;44(6):1107-1117). Another unpublished retrospective study showed that using CC Priming can get more oocytes with lower dose of GN. The study was conducted in our center with 131 subjects in CC group and 130 subjects in control group, in antagonist cycles.	clomiphene, but addition of clomiphene. Furthermore, as stated in the ESHRE guideline manual, "unpublished clinical trials should be avoided to support any recommendation".
6	Yuan Li	63	1580	Recommendation: "The addition of recombinant LH (rLH) to recombinant FSH (rFSH) is probably not recommended for controlled ovarian stimulation in the general IVF/ICSI population." Proposed change: "The addition of recombinant LH (rLH) to recombinant FSH (rFSH) is probably recommended for controlled ovarian stimulation in a subset of IVF/ICSI with LH deficiency." Evidence: Luteinizing hormone plays an essential physiologic role in follicle steroidogenesis and development and oocyte maturation (Curr Opin Obstet Gynecol,2002). Adequate LH level will benefit in embryo development. In a pilot dose-finding study in which patients were randomized to receive rhLH (0, 25, 75 or 225 IU/day) in addition to a fixed dose of rhFSH (150 IU/day), it was shown that a daily dose of 75 IU rLH was effective in most women in promoting optimal follicular development and enhancing the ability of these follicles to luteinize when exposed to hCG (J Clin Endocrinol Metab 1998; 83:1507-1514). An even more recent multi-center study (Hum Reprod 2001; 16:2525-2532) provided further evidence supporting this contention. The meta-analysis published in 2014 showed that Significantly higher clinical pregnancy rates were observed with r-hFSH plus r-hLH versus r-hFSH alone in the overall population analysed in this review (risk ratio [RR] 1.09; 95% CI 1.01–1.18) and in poor responders (n = 1179; RR 1.30; 95% CI 1.01–1.67; ITT population); the observed difference was more pronounced in poor responders. In total, 40 RCTs (6443 patients) were included in the analysis. Data on the number of oocytes retrieved were reported in 41 studies and imputed in two studies. Therefore, data were available from 43 studies (r-hFSH plus r-hLH, n = 3113; r-hFSH, n = 3228) in the intention-to-treat (ITT) population (all randomly allocated patients, including imputed data). Overall, no significant difference in the number of oocytes retrieved was found between the r-hFSH plus r-hLH and r-hFSH groups (weighted mean difference .0.03; 95% confidence interval [CI] 0.41 to 0.34). However, in poor responders, significantly more oocytes were retrieved with r-hFSH plus r-hLH versus r-hFSH alone (n =1077; weighted mean	The first two papers are in WHO group I anovulation patients, a patient population which is not in the scope of this guideline. We used the most recently published Cochrane meta-analysis to investigate the PICO question, which is by Mochtar et al. 2017. The Cochrane review and the thereafter published RCTs showed no benefit of LH supplementation neither in the general population, low responders or women of advanced age. After extensive discussion, the GDG could not reach consensus with regards to rFSH+rLH and decided not to formulate any recommendations on the topic.

				difference +0.75 oocytes; 95% [CI 0.14,1.36). According to our experience, there were 18% patients from RIF and 10% patients from general population have LH deficiency and need LH supplement. We found that using LH could improve the pregnant outcomes. Our study of Exome Sequencing showed LH deficiency patients were with Exome mutation. The results indicated there were still some LH deficiency patients exist in general population, who needs LH supplement.	
11	Yuan Li	89	2468	Recommendation: "The addition of a hormonal panel consisting of a combination of oestradiol, progesterone and LH measurements to ultrasound monitoring is probably not recommended." Proposed change: ""The addition of a hormonal panel consisting of a combination of oestradiol, progesterone and LH measurements to ultrasound monitoring is probably recommended." Evidence: LH monitoring is very important in protocol adjustment. It helps to optimize the usage of antagonist. In my previous published study (Frontiers in Endocrinology, 2019), we found that LH monitoring could help to improve the pregnancy outcomes in general population. A total of 567 women stimulated with recombinant FSH monotherapy in a GnRH antagonist protocol were studied. Among them, 256 patients showed relatively low LH levels [highest LH level (LHmax) <4 IU/L] during the entire ovarian stimulation process; 88 (Group A) and 168 patients (Group B) were stimulated without and with antagonist co-treatment, respectively. The remaining 311 patients had LHmax≥4 IU/L and were stimulated with a modified flexible antagonist protocol based on LH levels (Group C). The clinical and ongoing pregnancy rates were significantly higher in group A than group B (69.3 vs. 54.7%, P = 0.03 and 62.5 vs. 48.2%, P = 0.04, respectively), but the primary outcome measures did not differ between groups B and C. There were no significant differences in terms of patient demographics, LH levels, total dosage of gonadotrophin, duration of stimulation, follicular output rate between groups A and B, and between groups B and C. Also, there were no significant differences in laboratory and clinical outcomes in pairwise group comparisons. No canceled cycles due to premature ovulation was reported among the treated patients. This study indicated that LH levels may be used as an indicator for the time of antagonist addition. Patients with sustained low LH levels (LHmax<4 IU/L) during controlled ovarian stimulation (COS) might not require antagonist administration which may lead to low prognosis. Meanwhile, in patients with normal ovarian reservation, LH level monitoring is preferred to be used as the indicator of antagonist administration.	This study by Liu et al. 2019 may indeed be very interesting to optimize the usage of GnRH antagonist during ovarian stimulation, however, it does not compare monitoring of ovarian stimulation with either ultrasound or ultrasound and LH levels, which was the PICO question in this section. Your trial is eagerly awaited, and it will be incorporated in the update of this guideline.

				A well designed randomized controlled trials (RCTs) has been registered and initiated in our site to confirm our results, a novel treatment regimen based on LH measurements during COS might provide clinicians new insights about when to start antagonist administration in the GnRH antagonist protocol which could ensure the adequate LH level in follicular development and improve the outcome of COS treatment.	
9	Pedro Barri	80	2161	We recognize that there are no prospective randomized trials (RCT) that compare dual stimulation with two conventional stimulations in terms of efficacy (cumulative live birth rates) or efficiency (reduced time to live birth) of the two strategies. We also recognize that mandatory freeze-all of oocytes or embryos may be a disadvantage of this protocol because of additional procedure and oocyte manipulation, which, may not be allowed by some national health care. Nevertheless, it must be noted that freeze-all is mandatory also in case of luteal phase stimulation-only, random start, oocyte/embryo accumulation through sequential conventional stimulations and blastocyst stage PGT-A cycles. In addition, we do not understand why Committee Members did not mention in the evidence section that, according to all the papers published on the topic, the mean number of oocytes retrieved in the luteal phase stimulation is significantly higher than follicular phase as are the mean number of blastocysts and of euploid blastocysts. Moreover, the chance to find an euploid embryo or a blastocyst to transfer is significantly higher per started ovarian cycle in the dual stimulation if compared to standard stimulation (Ubaldi et al. 2016). While we understand that this procedure cannot be suggested for standard patients, poor prognosis patients (e.g. with reduced ovarian reserve, AMA, Bologna POR) might benefit from it. In conclusion, although there are no RCTs that show the superiority of dual stimulation vs conventional stimulation in terms of efficacy and efficiency, the author of this guideline could not undervalued the available evidence we believe that there are enough clinical data to state that "dual stimulation can be considered in poor prognosis patients when freeze-all is mandatory. It is not clear why the Committee stated that: "Luteal phase stimulation could be used in the nontransfer cycles" although it has far less clinical and laboratory evidence (some of which use data from dual stimulation) reported in the literature. Finally, dual stimulation is applied successfully by many centers in different countries. And the evidence published in favor of this procedure in increasing day by day (Xu and Li, 2013; Kuang et al., 2014; Moffat et al., 2014; Ubaldi et al, 2016; Wei Li-Hong et al.,	We have to look for the best evidence and as mentioned, there is no randomized trial (neither, considered in second hand, retrospective one) comparing dual stimulation with two conventional stimulations in term of the efficacy (cumulative live birth rates or at least number of oocytes) or efficiency (reduced time to live birth). The GDG agrees that in case of urgent fertility preservation (see this section), the dual stimulation, when it is possible, is the can be an option to get more eggs in a shorter time. However, in the dual stimulation there is a proportion of patient that can get an ongoing pregnancy in the first cycle and don't need the expense of a second cycle.

				2016; Tsampras et al., 2017; Vaiarelli et al, 2017; Cardoso et al., 2017; Liu et al., 2017; Cimadomo et al., 2017; Zhang Wei et al., 2018; Rashtian and Zhang 2018; Madani et al., 2018; Bailing Jin et al 2018; Vaiarelli et al 2018; Sighinolfi, Sunkara, La Marca, 2018; Alsbjerg, Humaidan et al., 2019).	
1	Fabiola Beligotti *	25	440	Recommend that a more recent systematic review by Iliodromiti et al. (Human Reproduction Update, Vol.21, No.6 pp. 698–710, 2015) should also be included in the references	This review by Iliodrometi et al. 2015 was excluded from the body of evidence because it is a systematic review without meta-analysis. Instead the IPD meta-analysis by Broer et al. 2013 was used.
1	Fabiola Beligotti *	27	478-482	Recommend that a more recent systematic review by Iliodromiti et al. (Human Reproduction Update, Vol.21, No.6 pp. 698–710, 2015) should also be included in the references	This review by Iliodrometi et al. 2015 was excluded from the body of evidence because it is a systematic review without meta-analysis. Instead the IPD meta-analysis by Broer et al. 2013 was used.
1	Fabiola Beligotti *	32	598-602	Automated methods for AMH measurement have demonstrated better precision than manual ELISA methods. Ref: van Helden et al (Human Reproduction 2015;30(8):1918–1926), Nelson et al (Fertil Steril 2015;104(4):1016–21), Fleming et al (Hum Fertil 2017 Jun 8:1-5), Li et al (Hum Reprod. 2016;31(12):2796-2802), Hyldgaard et al (Reprod Biol Endocrinol. 2015 Sep 22;13:107)	In the recommendation we do not address which type of assay should be used. Studies with all different assays could be included in the analysis. Depending on the assay that is used in a clinic assay-specific cut offs can be applied.
3	Fabiola Beligotti *	16	R 6	Evidence is mainly coming from trials in which rFSH only was applied (Griesinger et al. Fertil Steril 2010;94:2382-4), question is whether this effect is also applicable when hMG is applied.	This limitation is reported in the justification section of this recommendation.
3	Fabiola Beligotti *	15	R 5	Recommendation 5 is partially in contradiction with recommendation 6 (5: “The GDG acknowledges that oestrogen and progesterone are widely used for scheduling purposes. This is probably acceptable given the data on efficacy and safety.” – 6:” COCP pre-treatment (12-28 days) is not recommended in the GnRH antagonist protocol because of reduced efficacy)	In Recommendation 5: 'oestrogen and progesterone' was changed into 'oestrogen or progesterone' to make the difference with COCP clearer.
6	Fabiola Beligotti *	60	1503	In the Justification section it states that ‘The results from the meta-analysis suggest a slightly higher efficacy (LBR/PR) with hMG compared to rFSH in GnRH agonist cycles. However, the difference is not considered clinically relevant,’. There is no Global	According to the Cochrane meta-analysis, all gonadotropins are comparable in efficacy and with no

				consensus on clinically relevant difference in pregnancy rates and suggest the GDG avoids these interpretations	difference in safety, the GDG followed this conclusion.
14	Fabiola Beligotti *	22	R 61	Recommendation 61 does not fit the definition of a high ovarian response on page 97 which was more than 19 follicles or oocytes, it would be logical to align and be consistent.	Taking into account all the included studies into the guideline, not only diameter but also number of follicles are different. For example, in Griesinger' s study ≥ 19 but for Papanicolau ≥ 18 follicles, that could lead to incompatibility. Thus, we decided to put ≥18, also in the introduction.
6	Julia Koloda	18	12-14	«The addition of recombinant LH (rLH) to recombinant FSH (rFSH) is not recommended for controlled ovarian stimulation in low responders and women of advanced age». This statement seems to be controversial. The meta-analysis of Lehert, et al., 2014 showed a beneficial effect of rLH supplementation in poor-responders. A recent systematic review (Alviggi, et al., Fertil Steril. 2018 Apr;109(4):644-664) also demonstrated the benefits of recombinant hLH in two subgroups of patients: 1) women with adequate prestimulation ovarian reserve parameters and an unexpected hyporesponse to r-hFSH monotherapy; and 2) women 36-39 years of age.	The systematic review by Alviggi et al (2018) was not accompanied by a meta-analysis and thus it did not produce effect estimates that could be used in formulating a recommendation. The meta-analysis by Lehert et al (2014) was excluded for reasons stated in the guideline (see annex 7). After extensive discussion, the GDG could not reach consensus with regards to rFSH+rLH and decided not to formulate any recommendations on the topic.
16	Julia Koloda	23	6-8	«600 mg daily at least for micronized vaginal progesterone capsules and 300 mg daily at least for micronized vaginal progesterone suppositories/capsules». Please detail the difference between the drugs. It looks like the same preparations with differential doses are equally effective.	The GPP was adjusted and should now mention all available formulations.
4C	Julia Koloda	17	15-17	The statement «A higher gonadotropin dose of 300 IU is probably not recommended over the conventional dose of 150 IU for predicted low responders» is controversial to the comments «A higher gonadotropin dose of 300 IU daily results in a higher number of oocytes in low responders, and more chances of having an embryo for transfer»	The recommendation was changed to 'it is unclear whether a higher gonadotropin dose is recommended over 150 IU. Despite the higher number of oocytes, there was no difference in live birth/ongoing pregnancy rate. Furthermore, the

					sample sizes of the studies are small and therefore not sufficient to provide evidence for dose comparisons for live birth outcome.
	Ferraretti A.P, Gianaroli L., Fauser B., Tarlatzis B.	9	243	We can understand the rationale underlying the proposal of using the expressions "Low response" instead of "Poor response" but, in our opinion, before making this important modification to a well-assessed and generally accepted terminology, some clarifications should be made between terminology and definition.	The terminology has been adjusted in the guideline, from low responder to poor responder. However, the GDG thinks that 'poor response' can be potentially stigmatising/traumatising towards patients. Therefore, the GDG would like to propose to use the terminology low, normal and high response to categorize (predicted) response to OS for future referencing.
	Ferraretti A.P, Gianaroli L., Fauser B., Tarlatzis B.	9	253-256	"Low response" is defined as "≤ 3 follicles on day of oocyte maturation trigger and/or ≤ 3 oocytes obtained". ESHRE's current definition of POR was developed with the aim not only to define a "low response", but also to identify a sub-group of "patients" with a low response due to a "reduced ovarian reserve" and having the highest risk to be subject to reiterate poor ovarian responses with any stimulation strategy. To reach this goal and to identify a more homogeneous group of patients responding to this definition, reduced ovarian reserve should not only be "predicted" by means of ovarian reserve tests (10-15 % rate of false positive), but it should also be "validated ", or "confirmed", by the concomitant presence of an additional criteria. This specific sub-population may differ regarding the prognosis of pregnancy in relation to age, but the overall prognosis is always lower compared to similar age patients with a normal ovarian reserve. Professionals working in this field know very well that a "low response" as "generally" defined in line 253-256 may occur also in women with a normal ovarian reserve for intrinsic or extrinsic factors not related to the ovarian reserve. This second population of patients may significantly differ in many aspects from the one described in the previous paragraph and they might have different needs. It is clear for us that in point 4C , the recommendations are referring to the "Low responders" due to a reduced ovarian reserve "predicted" by AMH or AFC tests.	Due to the lack of universally accepted definitions of high and poor ovarian response, the definitions and terminology in the studies included in the evidence synthesis were varied.

				However, some concepts very clear for all of us may not be equally explicit for other readers.	
4C	Ferraretti A.P, Gianaroli L., Fauser B., Tarlatzis B.	48-53	1088-1285	With reference to POR patients classified according to the current ESHRE definition, it is true that none of the stimulation strategies tested was able to produce better results compared to any of the others. Therefore, it is likely that an "optimal stimulation" for this group of patients does not exists. However, currently available evidence (on which these recommendations are based) at least tells us which strategies should not certainly be used (high amount of Gn, natural cycles, rLH, additional drugs etc) or which one should be "used equally" (agonist or antagonist protocols etc). We would like to underline here that a relevant part of this current evidence derives from retrospective and prospective studies carried on based on the ESHRE definition of POR. [Just an additional observation (not so crucial!!) regarding "Expected" Low responders: ORTs (AMH or AFC) as a pre-treatment tool to choose the stimulation protocol should be used taking into consideration their false positive rate which is around 10-15% . In these cases, for example, an agonist protocol with 300 IU of FSH may be not the best choice!!] On the contrary, the second group (low response with normal ovarian reserve) may have a better ovarian response with different approaches and, subsequently, a better prognosis of pregnancy. Is it clear enough that this second group is not included in recommendations 4C? In 2011, ESHRE created a definition of POR (Poor Ovarian Response) that is almost universally accepted and that nowadays is routinely used in research and current practice when referring to the POR according to ESHRE definition. Although sometimes criticized and used with different cut-offs, in general the concept developed by ESHRE definition has now entered current knowledge: the "Bologna criteria" almost reached 1000 citations since their publication back in 2011. In addition, we would like to remember that the term Poor Ovarian Responders (POR) is included in the 2017 Glossary of ICMART (supported also by ESHRE) and it is also cited in The Practice Committee of the ASRM on "Testing and interpreting measures of ovarian reserve" (Fert Steril , Feb 2015) as the only definition available for poor ovarian responders. A new definition by the same Scientific Society based on a lexical choice rather than on a specific clinical concept might generate useless confusion and affect future studies on this subject, which are necessary to clarify still unsolved issues. To summarize, the term may be changed from Poor Ovarian Response (POR) to Low Ovarian Response (LOR), but it should be made clear for readers that this new terminology applies only to the POR patients according to the ESHRE definition.	The GDG is very aware that there are more nuanced definitions of a low ovarian response but decided to try and keep it simple: in the guideline there is only poor-normal-high responder. Therefore, the GDG only suggested definitions for low and high response.

				However, based on the above-mentioned issues, potential consequences of this change on clinical work, classification of patients and scientific research should be carefully taken into consideration. The acronym POR (Poor Ovarian Response) is almost universally accepted and nowadays it is routinely used in research and current practice when referring to the women identified by the ESHRE definition concepts. On the contrary, the new term Low Ovarian Response is not supported by any official document and, therefore, may refer only to "ovarian response" and not to a "homogeneous group of women". In conclusion, although sometimes criticized and used with different cut-offs, the concept of the ESHRE definition of POR has entered current knowledge. Therefore, it would be senseless that the same scientific community reduces the power of what has been achieved in eight years of hard work by modifying a useful definition based on these premises. For all these reasons, our final request is do not change the term from Poor to Low.	
11	Ferraretti A.P.	89	2459-2478	You never mention the progesterone measurement on the day of triggering. As you know, high levels (cut-off not defined!) are often considered in the practice as a criterium for freeze- all. If the GDG agree that the current evidence do not support this strategy, I believe that it would be very important to dedicate a specific recommendation to avoid useless (and over-treatment) freeze- all strategies.	This question was not within the original questions of this guideline. This will be considered to be included in the update of the guideline.
6	Ferraretti A.P.	62	1567	In the paper I published in 2014 in POR, the rLH was not utilized in addition to FSH but as a pre-treatment in agonist protocol with the aim to increase the intraovarian levels of androgens before the FSH administration. A scientific background exists, in my opinion, to test this approach in POR (Weil et al,1998,1999, Hillier et al 1981,1997, Vendola et al, 1998) : - accumulation of androgens in the micro milieu of the primate ovary plays a crucial role in early follicular development and GC proliferation, - androgens increase stimulate early stage of follicular growth and increase the number of preantral and antral follicles . - increased intraovarian concentration of androgens augments FSH receptor expression in GC and thus, potentially lead to enhanced responsiveness of ovaries to FSH, Most protocols based on this concept are using a pre-treatment with exogenous androgens. The limitation of their efficacy may be that they do not reach the intraovarian milieu; in addition, by suppressing the endogenous LH, they may also reduce the intraovarian production of androgens. On the contrary, exogenous LH may	This was adjusted in the guideline.

				increase the intraovarian production of androgen. So, I ask you to change the term “supplementation” with “pre-treatment”	
17	Mochtar M., van Wely M., Braat D., Goddijn M.	116-117	3266-3303	In current draft Guideline, there is no mentioning of GnRH agonist trigger failure at all, although we are all aware that GnRH agonist trigger failure exists in approximately 1-2% of cycles, resulting in a serious failure to recover oocytes, and once it happens it has an enormous impact on the women/couple under concern. Can GnRH agonist trigger failure and its potential rescue policies (e.g. hCG retrigger/ LH measurements) be mentioned in short in the Guideline in chapter 17.1? (see a.o. Chang, FS 2016)	This question was not within the original questions of this guideline. This will be considered to be included in the update of the guideline. Furthermore, this topic is going to be discussed in the fertility preservation guideline
10	Mochtar M., van Wely M., Braat D., Goddijn M.	84-85	2298-2347	In view of the lack of evidence of a beneficial effect to the addition of anti oestrogens/ aromatase inhibitors for ovarian hyperstimulation in women with breast cancer (no trial outcomes worldwide hitherto), we fear that any recommendation introducing the advice to add one of these agents will be a barrier for recruitment of patients in ongoing trials and may provoke implementation of the wrong recommendations. On top of that, eventual later de-implementation will be extremely difficult. FYI: Results of the STIM trial will be presented at the October ASRM 2019 meeting. Preliminary -not yet peer reviewed data- show no effect of the concomitant use of Letrozole of Tamoxifen on oocyte yield. We are fully aware that preliminary data can’t be part of the guideline, but ask for a more evidence based approach of the recommendation of concern. Can the wording for the recommendation in chapter 10.3 “In ovarian stimulation for fertility preservation in oestrogen sensitive diseases the concomitant use of anti-oestrogen therapy, such as letrozole or tamoxifen, is probably recommended” be changed into : “Whether In ovarian stimulation for fertility preservation in oestrogen sensitive diseases the concomitant use of anti-oestrogen therapy, such as letrozole or tamoxifen is beneficial, should be evaluated in the context of ongoing trials’	The recommendation was changed into a GPP and made less strong.
10	Richard Anderson et al.	83	R49	As there is also a general strong recommendation stating “The GnRH antagonist protocol is recommended over the GnRH agonist protocols given the comparable efficacy and higher safety in the general IVF/ICSI population (recommendation 24 - strong, ⊕⊕⊕O), we do not see the reasoning why the strongly recommended safer option is only ‘probably recommended” in fertility preservation patient, which could be considered a more fragile IVF population. We would like to suggest changing this recommendation to a strong recommendation for the safest option, based on	The GDG agrees with the comment, however, the general IVF population and a fertility preservation population are very different. These are not always women with a high ovarian reserve, and the necessity of oocyte vitrification may urge the

				extrapolated evidence from the general IVF/ICSI population. In addition, we think the justification should state that Data on live births are limited, in particular in cancer patients having vitrified oocytes. (add reference cobo dec 2018)	need to use GnRH agonist protocols for planning purposes.
10	Richard Anderson et al.	84	R50	For patient undergoing urgent OS for Fertility preservation, random-start is the first line option. Therefore, we could like to ask to change this recommendation to "random-start OS is an important option" in the recommendation.	The recommendation was altered as requested.